

2009

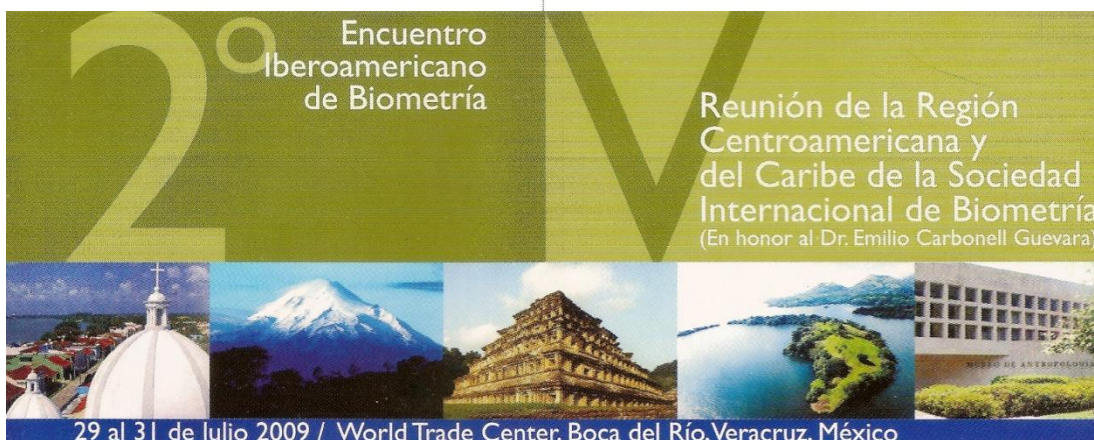
BioS³ metria



The International
Biometric Society



Universidad Veracruzana



Memoria del 2° Encuentro de Biometría y la V Reunión de la Región Centroamericana y del Caribe de la Sociedad de Biometría

Sergio Francisco Juárez Cerrillo
Mario Miguel Ojeda Ramírez
Editores

www.uv.mx/eib

Memoria del 2° Encuentro Iberoamericano de
Biometría y la V Reunión de la Región
Centroamericana y del Caribe de la
Sociedad Internacional
de Biometría

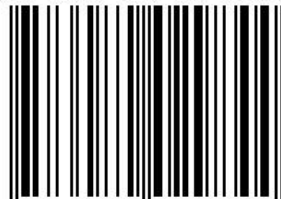
Sergio Francisco Juárez Cerrillo
Mario Miguel Ojeda Ramírez

Editores

Xalapa, Ver., México

Mayo 2010

ISBN 978-607-00-2984-4



9 786070 029844 >



*Proyecto realizado con financiamiento de la Secretaria de Educación Pública –
Subsecretaria de Educación Superior – Dirección General de Educación
Superior Universitaria, Convenio No. 2009-30-001-032*



2^o Encuentro Iberoamericano de Biometría

Bioestadística

V Reunión de la Región Centroamericana y del Caribe de la Sociedad Internacional de Biometría (En honor al Dr. Emilio Carbonell Guzmán)



Universidad Veracruzana

Julio 2009 Boca dl Rio,Ver



Sociedad Internacional de Biometría

foto: imagen_makiz@yahoo.com
Cel:0152 54856791, Cel: 0445 5514293302

CONTENIDO

| | |
|---|------|
| PREFACIO..... | XIII |
| Un Enfoque Bayesiano para el Problema de Colinealidad en Regresión Logística Francisco Julián Ariza Hernandez, Flaviano Godínez Jaimes, Ramón Reyes Carreto, Esperanza Nicolás Popoca | 1 |
| Un Estudio Comparativo de Estimadores de Índices de Diversidad: Una Corrección al Índice de Fager C. Bouza, Di Rienzo, D. Covarrubias, L. Pimentel | 7 |
| Comparación de la Arquitectura de dos Especies Simpátricas de Cactáceas Columnares <i>Cephalocereus palmeri</i> var. <i>sartorianus</i> y <i>Neobuxbaumia scoparia</i> en el Bosque Seco Caducifolio de Cerro Gordo, Veracruz Judith Guadalupe Montero Mora, Armando J. Martínez, Sergio Francisco Juárez Cerrillo | 11 |
| Un Modelo de Regresión Poisson Inflado con Ceros para Analizar Datos de un Experimento de Fungicidas en Jitomate María de Lourdes Velasco Vázquez, Sergio Francisco Juárez Cerrillo | 16 |
| Una Propuesta para Obtener la Estimación Máximo Verosímil cuando la Ecuación de Verosimilitud Presenta una Estructura Compleja Ernesto Menéndez Acuña | 22 |
| Sobre la Potencia Prospectiva: un Programa para su Cálculo Luis Omar Jamed, Ernesto Menéndez Acuña | 27 |
| Contribución de la Psicología en el Desarrollo de la Estadística Mtro. José García Franco..... | 32 |
| Actitud hacia la Estadística en Estudiantes de Estadística e Informática José Luis Colorado Hernández, Sergio Francisco Juárez Cerrillo..... | 37 |
| Propuesta de una Escala Alternativa para Medir Hábitos de Estudio en el Universitario Ma. de los Ángeles Silva Mar, María de los Ángeles García Oramas, Claudio Rafael Castro López | 43 |
| Interpretación y Construcción de Gráficas: Dificultades del Estudiante Egresado del Bachillerato Abraham Cuesta Borges, Marco Antonio Méndez Salazar..... | 49 |
| Entornos Geométrico y Algebraico: Dificultades del Estudiante Egresado del Bachillerato Abraham Cuesta Borges, Marco Antonio Méndez Salazar..... | 54 |
| Ocurrencia Temporal de Parásitos en <i>Macrobrachium acanthurus</i> de la Cuenca Baja del Río Jamapa Magda Estela Domínguez-Machín, Sergio Francisco Juárez Cerrillo | 59 |
| Impacto del Estrés Materno, Exposición a Plomo, Anemia por Deficiencia de Hierro Pre y Postnatal en el Desarrollo Neuroconductual Infantil Siliceo-Murrieta, JI., Lamadrid-Figueroa, HM., Trejo-Valdivia, B., Mercado-García, A., Wright, RO., Téllez-Rojo, MM., Wright, R, Lourdes Schnaas | 64 |
| Estimación de Estadísticos de Primer y Segundo Orden Basados en el Análisis de Imágenes Digitales de la Naranja Valencia Barradas-Domínguez P., Jiménez-Fernández M., Atanasio-Gómez R. | 70 |
| Funciones de Varianza, Covarianza y Correlación Bicuadrática para Distribuciones Normales Carlos Eduardo Alonso Malaver, Jorge Martínez Collantes..... | 75 |
| Inferencia Sobre Datos No Detectados de Contaminantes Dr. Fidel Ulín Montejó, Dr. Humberto Vaquera Huerta..... | 80 |
| Modelos de Regresión Múltiples para Explorar el Contexto Paisajístico sobre la Biodiversidad en un Agropaisaje Mesoamericano S. Vílchez, CA. Harvey; F. Casanoves, B. Finegan, J. Saenz, D. Sánchez, A. Medina, Hernández | 87 |
| Diagnóstico de Dengue: Una Aplicación de Regresión Logística Binaria. Ana María Aguilera, Agustín Santiago..... | 93 |
| Algunos Modelos Epidemiológicos Estocásticos Cruz-Kuri, Luis, Sosa-Galindo, Ismael, García-Banda, Agustín J..... | 98 |
| Análisis Multinivel: Una Generalización del Análisis de Covarianza de Efectos Aleatorios Wilmer Fermín, Haidee Castañeda..... | 103 |
| Representación del BLUP en el Modelo Lineal General Mixto, Caso Balanceado Fernando Velasco Luna, Mario Miguel Ojeda Ramirez | 108 |
| Evaluación de Insectidas para el Control de Garrapatas (<i>Boophilus microplus</i>) que Afectan al Ganado Bovino de Doble Propósito usando Modelos Lineales Generalizados Rafael Borges, Andrea González, Miguel Pérez, Mandius Carvajalino, Daira Velandia, Diliana Tapias | 114 |

| | |
|---|-----|
| Uso de los Modelos Monofásico, Difásico y Gamma Incompleta para Caracterizar Curvas de Lactancia en Cabras Lecheras Castañeda-Bustos Vielka Jeanethe, Torres-Hernández Glafiro, Hernández-Mendo Omar, Mandeville Peter B., Díaz-Gómez Marta Olivia, Ochoa-Cordero Manuel Antonio, Pérez-Elizalde Sergio, González-Camacho Juan Manuel..... | 119 |
| Algunos Aspectos Estadísticos sobre un Estudio de Salud Nutricional y Salud Dental en Niños de Enseñanza Media en el Estado de Veracruz Luisa Otilia Navarro Ortiz, Luis Cruz-Kuri, Paula Yadira Landa Cano, Julia Leticia Beltrán Martínez, María Magdalena Palmeros Rodríguez | 124 |
| Algunas Estadísticas de la Precipitación en el Puerto de Veracruz Antonio Luna Díaz Peón, Sergio Hernández González | 129 |
| Propuesta, y Ejemplo, de un Procedimiento Multivariante para Integrar la Metodología Cualitativa (con Material de Texto Libre) y la Metodología Cuantitativa (Estadística Multivariante) José David Urchaga Litago..... | 135 |
| Simplificación de los Análisis Morfométricos para Poblaciones de la Tortuga Dulceacuícola kinosternon herrerai Gustavo Aguirre León, Erasmo Cázares Hernández, Blanca Hervier Martínez | 140 |
| Estudio Multivariante de las Características del Turista Español que Visita México Jaime Egido, Purificación Galindo | 145 |
| Regresión Logística Ordinal: Aplicación en Niveles de Estrés Luis F. Grajales, Iván Mendivelso..... | 151 |
| Descripción estadística de la evaluación clínica y neurofisiológica del Síndrome de Guillain-Barré en México Juan José Méndez Castillo, Luis Angel Ruano Calderón, Mariangeles Moreno Barjollo, Lucas López Segovia..... | 156 |
| Factores no Organizacionales y su Posible Relación con Burnout en Personal de Enfermería de Unidad Coronaria Virginia Cortés Córdova, Jorge Elías Torres López, Lucas López Segovia..... | 161 |
| Estudio Inmunohistoquímico de los Variables Categóricas Ordinales por Caspasa 3 y Alfa-Fetoproteína de las Lesiones Hepáticas Producidas por el Virus de la Hepatitis C y su Relación con el Hepatocarcinoma. Captación Genómica por Fusión Celular: Una Teoría de la Involución de las Especies Leonardo Varela Rodríguez, Martín Varela Vindas | 167 |
| Estudio Inmunohistoquímico de los Variables Categóricas Ordinales por P53 y Alfa-Fetoproteína de las Lesiones Hepáticas Producidas por el Virus de la Hepatitis C (VHC) y su Relación con el Hepatocarcinoma (HCA) Leonardo Varela Rodríguez, Martín Varela Vindas | 174 |
| La Rugoscopía Palatina Humana como Factor Fundamental de Identificación Personal Noralia Ramírez Chávez, Alfonso Sotelo y Soto Gustavo, Jorge de Jesús Serena Melo, Guillermo Hernández Lira, Rafael Castillo Pratz, y Rafael Alonso Nuñez..... | 180 |
| Aportaciones a la Medida de la Fiabilidad del Instrumento Purpose in Life (PIL) Dr. José David Urchaga Litago, Lic. Eliseudo Salvino Gomes, y Dr. Antonio Sánchez Cabaco | 186 |
| Sistema para Evaluar el Grado de Percepción de los Clientes sobre una Organización Prestadora de Servicios Juan de Dios Villaseca Galván | 191 |
| Modelos Matemáticos para la Estimación del Desarrollo de Novillas: Edad y Peso a Primera Concepción Roberto Castillo Tlapa, Belisario Domínguez Mancera, Antonio Hernández Beltrán, José Manuel Martínez Hernández, Patricia Cervantes Acosta, Blanca Catalina Colin Ibarra..... | 196 |
| Estudio del Bajo Peso en Escolares de 6-13 Años en 7 Escuelas de la Ciudad de Chilpancingo, Gro., México Vivian Sistachs Vega, Lucio Díaz González | 201 |
| Inventario de Maguey Papalote (Agave cupreata Trel et Berger) de la Región Centro del Estado de Guerrero. L. Alonso, D. Covarrubias, F. Maradiaga..... | 205 |
| Estudio Comparativo entre Algoritmos de Clasificación Automática usando Simulaciones con el Paquete WEKA Capilla, J., Reyes, H., Linares, G..... | 209 |
| Evaluación del Efecto del Tratamiento Térmico sobre los Carotenoides en Galletas Adicionadas con Espinaca Mediante Superficie de Respuesta A. Reyes-Téllez, E.N. Aquino Bolaños, R.I. Guzmán, I. Verdalet Guzmán, D. Leyva Daniel, M. Vázquez Hernández, Z.J. Dominguez Esquivel, E.R. Silva Hernández | 213 |

| | |
|--|-----|
| Estudio de la Liberación de Sabor en Goma de Mascar, Utilizando Microcápsulas de Aceite Esencial de Menta S. C. Pereyra, A. Cano, M. Jiménez | 218 |
| Influencia de la Concentración de Almidón de Maíz y Mucilago de Nopal (<i>Opuntia ficus indica</i>) en la Caracterización Mecánica de Películas Extrudidas Comestibles Utilizando el Análisis de Superficie de Respuesta Morín-Alcázar C.S., Calvo-López A., Galicia-García T., Martínez-Bustos F., Grosso C., Aguilar-Palazuelos E. | 223 |
| Efecto de la Concentración de Almidón de Haba (<i>Vicia faba</i> L) en las Propiedades Mecánicas de Películas Comestibles. Calvo-López A., Morín-Alcázar C.S., Galicia-García T., Martínez-Bustos F. | 228 |
| Propiedades Mecánicas en Láminas para Bolsas Biodegradables Obtenidas por el Proceso de Extrusión Termoplástica. Análisis de Superficie de Respuesta Galicia-García, T., Martínez-Bustos, F., Jiménez-Arévalo, O. | 233 |
| Uso del Modelo de Mezclas para Determinar la Asociación Óptima de Gramíneas-Leguminosa Rigoberto Castro Rivera, Alfonso Hernández Garay, Humberto Vaquera Huerta, Jorge Pérez Pérez, José de la Paz Hernández Girón, Adrian Raymundo Quero Carrillo, Omar Ramírez Reynoso, Javier Francisco Enríquez Quiroz..... | 238 |
| Una actividad conducente a lograr el aprendizaje significativo del concepto de error tipo I Félix Almendra-Arao, David Sotres-Ramos, Cecilia Ramírez Figueroa | 243 |
| Una experiencia Docente para el Aprendizaje de Modelos Probabilísticos Discretos Basadas en la Redacción de Enunciados de Problemas M. J. García-Ligero, P. Román-Román | 248 |
| Métodos de apoyo/recursos interactivos de aprendizaje de SPSS Lara-Porras, A.M., Alfonso-Uxó, A., Jiménez-Lara Pablo..... | 253 |
| Recursos didácticos para el aprendizaje de la probabilidad: Una experiencia virtual Lara-Porras, A.M., Román-Montoya, Y. | 259 |
| Análisis Exploratorios de Hábitos de Alimentación y Actividad Física en Escolares de la Ciudad de Chilpancingo del Estado de Guerrero México Vivian Sistachs Vega, Lucio Díaz González | 265 |
| Análisis Estadístico de Datos Textuales M. Luisa Hdez. Maldonado, Claudio R. Castro López | 270 |
| A Consideration for Constructing Critical Regions for Non-inferiority Tests Dr. Félix Almendra-Arao..... | 275 |
| La Explicación del Funcionamiento de un Sistema Mediante la Modelización de Ecuaciones Estructurales M ^a Elena Vicente, J. Antonio Castro, M ^a Purificación Vicente, M ^a Purificación Galindo..... | 280 |
| Modelos para Enfermedades Infecciosas Vivian Sistachs Vega, Lucio Díaz González, Lucio Díaz González..... | 285 |
| Estructura Causa Suficiente en Modelos Gráficos para Enfermedades Complejas con Expresiones Orales Mabel Brunotto, Silvia Mina, Magdalena Bolesina, Silvia Sorokin, Irma Moncunill, Susana Cornejo, Vanesa Buffa, Dante Secchi, Eduardo Piemonte, Evelin Bachmeier, Jerónimo Lazos, María Brezzo, Ana María Zarate, Fabiana Carletto Köber, Alicia Malberti, Silvina Barembaum, Ana Isabel Azcurra, José Luis Barra..... | 289 |
| Ajuste, Predicción y Tiempos de Primer Paso en un Modelo Estocástico para el Estudio de Crecimiento del Pez Espada Román-Román, Patricia, Torres-Ruiz, Francisco de Asís..... | 294 |
| Uso de Diversas Técnicas Estadísticas en el Diseño de una Bebida con Actividad Funcional Evangelina Ruiz-Figueroa, Elia Nora Aquino-Bolaños, Rosa Isela Guzmán-Jiménez, Diana Elizabeth Leyva-Daniel, Takuo Nakano, Eryck R. Silva-Hernández | 299 |
| Aplicación del Algoritmo Chaid para el Análisis de la Televisión y el Comportamiento en Niños Preescolares Laura Oliva Zárate, Claudio Castro | 304 |
| Comparando el Análisis de Componentes Principales (PCA), el Test No Paramétrico de Friedman-Nemenyi (TFN) y el Análisis Discriminante (AD) para la Clasificación de Expresiones Faciales en Bebés Cuando Toman una Fórmula Infantil Rosa Guadalupe Herrera-Lee, Iñigo Verdalt-Guzmán, Takuo Nakano, Lech Ozimek, María Luz Márquez-Barradas, Homero Vladimir Ríos-Figueroa, Jesús Ofelia Angulo-Guerrero, Eryck R. Silva-Hernández..... | 309 |
| Análisis de Parámetros Físicoquímicos, Color, Contenido de Polifenoles, Antocianinas y Capacidad Antioxidante en Licores de Mora mediante Técnicas Multivariadas | |

| | |
|--|-----|
| Diana Elizabeth Leyva-Daniel, Iñigo Verdalt-Guzmán, Eryck R. Silva-Hernández, Elia Nora Aquino Bolaños, Carlos Guillermo Hernández, Rosa Isela Guzmán-Gerónimo | 314 |
| Análisis de Interacción de Factores Clasificatorios de una Población en Base a las Distancias Moleculares entre sus Individuos | |
| Bruno, Cecilia, Balzarini, Mónica | 319 |
| Relación del Ángulo Interdental con el Plano Mento-Nasal (Plano Estético de Ricketts) | |
| Gustavo Alfonso Sotelo y Soto, Noralia Ramírez Chávez, Erika Sotelo Acosta, Juan José Trujillo Fandiño, Julia Aguirre Sampieri, Jorge de Jesús Serena Melo, Minerva Reyes Felix..... | 325 |
| Alometría Interespecífica de Árboles de la Selva Alta Perennifolia en el Volcán San Martín Pajapan, Veracruz | |
| Báez-Hernández. A., A. J. Martínez., S.M. Vázquez–Torres & J. G. Montero–Mora..... | 330 |
| Comparación de Estimadores de Regresión del Total en Poblaciones Estratificados y Complejos Bajo Tres Propuestas de la Matriz de Varianzas | |
| Flaviano Godínez Jaimes, Ignacio Méndez Ramírez, Ma. Natividad Nava Hernández..... | 336 |
| Un Estudio de la Evolución de Epidemias de Influenza en Estados Unidos y la Posible Asociación con el Factor Climatológico | |
| Agustin Jaime Garcia Banda, Luis Cruz-Kuri, Ismael Sosa Galindo..... | 341 |
| Algunos Modelos Estadísticos Simples para Describir la Evolución de un Proceso Epidemiológico | |
| Ismael Sosa Galindo, Luis Cruz-Kuri, Agustin Jaime Garcia Banda..... | 346 |
| Estudio Morfométrico de la Tortuga Kinosternon Herrerai en un Arrollo Situado en el Municipio de Xalapa, Ver. | |
| Gustavo Aguirre León, Erasmo Cázares Hernández, Blanca Hervier Martínez | 351 |
| Evaluación Sensorial de Aderezos Ricos en Ácidos Grasos Omega-3 y Omega-6 | |
| Martínez-Alejo J.M., García-Barradas O., Mendoza-López M. R., Muñoz-Muñiz O.D., Jiménez-Fernández M. | 356 |
| Comparación de Filtros Específicos y no Específicos en el Análisis de Matrices de Expresión Génica | |
| María del Carmen Romero, Julio Alejandro Di Rienzo..... | 362 |
| Utilización del Análisis de Cluster con Variables Mixtas en la Identificación Taxonómica de Ácaros Depredadores (acari phytoseiidae) de Cuba | |
| Dunia Chávez Esponda, Ileana Miranda Cabrera, Mayra Ramos Lima..... | 367 |
| Un Modelo SARIMA para Viento en la Estación Nucleoeléctrica de Laguna Verde | |
| Uziel González González, Alejandro Jiménez Lagunes, Sergio F. Juárez Cerrillo..... | 370 |
| Análisis Espectral Singular de Series de Tiempo | |
| Alejandro Ramírez Núñez, Sergio Francisco Juárez Cerrillo | 375 |
| Análisis Estadístico de la Implementación del Laberinto Acuático de Morris en la Evaluación del Aprendizaje Espacial | |
| Pérez Roldan E.A, Santiago Roque I, Juárez Aguilar E, y Juárez Cerrillo S.F. | 379 |

Prefacio

Varios académicos de la Universidad Veracruzana, y otros tantos de México, asistimos a Salamanca, España, al Primer Encuentro Iberoamericano de Biometría. Este evento se desarrolló en el marco de la XI Conferencia Española que la Región de la Internacional Biometric Society (IBS) celebrada en ese país los días 20, 21 y 22 de junio del 2007. En esa ocasión fuimos partícipes de un magno evento y una magnífica organización.

En este contexto se nos invitó a organizar el Segundo Encuentro Iberoamericano de Biometría y la V Reunión de la Región Centroamericana y del Caribe de la Sociedad Internacional de Biometría para el 2009. Convocamos al Grupo México, a la Región Centroamericana y del Caribe, ambos de la IBS, y a un conjunto amplio de instituciones de México y de otros países de Iberoamérica, y también tuvimos el respaldo de colegas de otras regiones, grupos y países. La sede del evento fue la ciudad de Boca del Río, y se llevó a cabo del 29 al 31 de julio.

El evento fue una oportunidad para constatar el avance de la Biometría en Iberoamérica y también permitió fortalecer los vínculos y las redes de colaboración de académicos y usuarios de esta importante disciplina. Cabe destacar que el Segundo Encuentro Iberoamericano de Biometría se organizó también para darle continuidad programática a la Red Iberoamericana de Biometría, actividad que llevan a cabo de forma conjunta las regiones Española, Argentina y Centroamericana y del Caribe y el Grupo de Chile de la IBS.

La V Reunión Centroamericana y del Caribe de IBS, forma parte de la programación bianual de la Región que, desde su cambio de estatus como región de la IBS, ha organizado las siguientes reuniones: en el Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) en Turrialba, Costa Rica en el 2001; en la Universidad de Puerto Rico en Mayagüez, Puerto Rico en el 2003; en el instituto de Ciencia Animal (ICA) en la Habana, Cuba en el 2005; y la más reciente, en la Universidad de Oriente-Núcleo de Nueva Esparta, en Isla Margarita, Venezuela, en el 2007.

En esta edición del evento se rindió homenaje al Dr. Emilio Carbonell Guevara, Jefe de la Unidad de Biometría del Instituto Valenciano de Instituciones Agrarias de Valencia, España y ex presidente de la Región Española de IBS, por su valioso aporte y por haber sido pieza fundamental en el desarrollo de la Región. También se le nombró miembro honorario de la Región Centroamericana y del Caribe.

La Biometría es una disciplina de fuerte desarrollo y amplia influencia en las más diversas actividades humanas; como tal las líneas de investigación teórica y aplicada se han multiplicado en los años recientes. Este segundo Encuentro Iberoamericano de Biometría se dedicó a cuatro temas fundamentales: Modelación Estadística en la Biometría, Análisis Multidimensional, Educación Estadística y Celebración de los 150 años de la aparición del libro “El Origen de la Especies” de Charles Darwin.

Lo que aquí presentamos son una parte de las solicitudes que, como contribuciones libres, recibió el Comité Científico por parte de la comunidad biométrica internacional. Los editores realizamos la revisión y adecuación de forma de los trabajos en extenso, de acuerdo a lineamientos difundidos en la convocatoria. En esta memoria ponemos a disposición las contribuciones libres editadas. Queremos manifestar nuestro más profundo agradecimiento a los autores.

Agradecemos también a nuestros patrocinadores por sus generosas aportaciones, las cuales fueron vitales para hacer realidad el evento. En particular agradecemos a Subsecretaría de Educación Superior, al Consejo Nacional de Ciencia y Tecnología, la Universidad Veracruzana, a la Secretaría de Educación de Veracruz, al Instituto de Ecología, a la Universidad de Colima y al Gobierno del Estado de Veracruz.

Finalmente agradecemos de manera muy especial a Rubén Jiménez Zamudio y Cecilia Cruz López por todo el apoyo en la edición de este documento.

Febrero de 2010.
Sergio Francisco Juárez Cerrillo
Mario Miguel Ojeda Ramírez
Editores.

Un Enfoque Bayesiano para el Problema de Colinealidad en Regresión Logística

Francisco Julián Ariza Hernández¹, Flaviano Godínez Jaimes², Ramón Reyes Carreto³, Esperanza Nicolás Popoca⁴

¹ Universidad Autónoma de Guerrero. fjah@colpos.mx

² Universidad Autónoma de Guerrero. fgodinezj@gmail.com

³ Universidad Autónoma de Guerrero. rcarreto1@yahoo.com

⁴ Universidad Autónoma de Guerrero

Resumen

Típicamente el problema de colinealidad en las variables explicativas provoca problemas en la estimación de parámetros sobre todo cuando se usan métodos convencionales como el de máxima verosimilitud. El caso de la regresión logística no es la excepción; además, en este tipo de análisis puede ocurrir separación o cuasi separación en los datos lo cual origina que el estimador de máxima verosimilitud no exista. En este trabajo, se presenta un enfoque bayesiano utilizando distribuciones aprioris no informativas al problema de colinealidad en las variables explicativas y separación en los datos en el modelo de regresión logística. Para obtener muestras de la distribución a posteriori se usa el algoritmo de Metrópolis. La comparación de los estimadores estudiados y otros encontrados en la literatura se realiza mediante simulación en términos de sesgo y error cuadrático medio.

Palabras clave: separación en los datos, colinealidad, estimación bayesiana, apriori no informativa.

1. Introducción

En muchas aplicaciones del modelo de regresión logística, hay situaciones en que la colinealidad es inevitable por ejemplo cuando varias variables son medidas en el mismo individuo, las cuales frecuentemente se encuentran de alguna manera relacionadas, o bien cuando se usa un número grande de variables comparada con el número de observaciones. Otro problema que afecta al modelo de regresión logística es la separación en los datos. Tanto separación como colinealidad causan que tanto los parámetros y las varianzas estimadas sean grandes lo que produce

que los intervalos de confianza asintóticos sean muy amplios, reduciendo así su utilidad práctica y que las pruebas de hipótesis pierdan potencia. Algunos estimadores que no son afectados por la existencia de separación en los datos se han basado en una función de penalty a priori no informativa de Jeffrey propuesta por Firth (1993) y retomado en el trabajo de Heinze y Schemper (2002). Rousseeuw y Christmann (2003) usan una codificación del modelo que denominan modelo de regresión logística escondida y Gelman *et al.* (2008) usan un enfoque bayesiano. Algunas soluciones al problema de colinealidad son estimación ridge (Schaefer *et al.*, 1984; Lee y Silvapulle, 1988; le Cessie y van Houwelingen, 1992; Godínez y Ramírez, 2005, 2006), la estimación en componentes principales y estimadores tipo Stein (Schaefer, 1986; Ramírez, 1997). Pocos trabajos han considerado el efecto conjunto de colinealidad y separación en el modelo de regresión logística. Godínez y Valverde, 2005, 2006 proponen estimadores ridge iterativos y de un paso mientras que Shen y Gao (2008) proponen una doble penalización de tipo a priori no informativa de Jeffrey y una penalización de tipo ridge. En este trabajo se presentan estimadores de tipo bayesiano basado en dos distribuciones a priori no informativas cuando hay separación y colinealidad en el modelo de regresión logística.

Sea $\{(Y_i, \mathbf{x}_i^t); i = 1, \dots, n\}$ una muestra en regresión binaria, esto es, Y_i son variables aleatorias independientes con distribución Bernoulli con probabilidad de éxito desconocida $\pi_i = P(Y_i = 1)$ y $\mathbf{x}_i^t = (x_{i1}, \dots, x_{ip})$ son vectores no estocásticos de variables explicativas de dimensión p ($p < n$). Sea \mathbf{X} la matriz diseño de $n \times (p + 1)$ cuyos renglones son $(1, \mathbf{x}_i^t); i = 1, \dots, n$. En regresión logística se supone que

$$\pi_i = P(Y_i = 1) = \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}_i^t \boldsymbol{\beta}})} \quad (1)$$

donde $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^t$ es el vector de parámetros desconocidos y $\mathbf{x}_i^t \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$; $i = 1, \dots, n$ como una combinación lineal de las variables que explican la probabilidad de éxito.

La función de verosimilitud del modelo de regresión logística se expresa como

$$\begin{aligned} L(\boldsymbol{\beta}; \mathbf{y}) &= \prod_{i=1}^n \{(\pi_i)^{y_i} (1 - \pi_i)^{(1-y_i)}\} \\ &= \exp \left\{ \sum_{i=1}^n y_i \mathbf{x}_i^t \boldsymbol{\beta} - \sum_{i=1}^n \log(1 + e^{\mathbf{x}_i^t \boldsymbol{\beta}}) \right\} \end{aligned} \quad (2)$$

2. Estimación Bayesiana

Se desea obtener estimadores Bayesianos de los parámetros β 's cuando hay presencia de colinealidad en las variables explicativas y de separación en las observaciones. Gelman *et al.* (2008) abordan este problema utilizando distribuciones a priori simétricas independientes para los parámetros tales como la t de Student, la normal y la distribución Cauchy con parámetros de localidad 0 y de escala 2.5. El proceso de estimación lo llevan a cabo usando el algoritmo EM (algoritmo de Esperanza Maximización) en el método de mínimos cuadrados ponderados iterativos. Los métodos y procedimientos usados por Gelman *et al.* (2008) están implementados en el programa estadístico R a través del comando `bayesglm` que se obtienen de la librería `arm`.

En este trabajo, se utilizan otras distribuciones a priori a las propuestas por Gelman *et al.* (2008). Ellas son, la distribución a priori no informativa y la distribución normal con varianza "grande", que en lo sucesivo llamaremos normal plana. Para obtener los estimadores se consideran las funciones de pérdida, pérdida cuadrática y la pérdida lineal absoluta.

Considere las siguientes distribuciones conjuntas a priori de los parámetros,

$$\begin{aligned} p_{NP}(\beta) &= \frac{1}{(\sqrt{2\pi})^{p+1}} |\mathbf{B}_0|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2}(\beta - \mathbf{b}_0)^t \mathbf{B}_0^{-1} (\beta - \mathbf{b}_0) \right\}; \\ p_{NI}(\beta) &\propto 1 \end{aligned}$$

donde $\beta_0 \in \mathbb{R}^{p+1}$, y los elementos de matriz \mathbf{B}_0 son tales que $(b_{ij}) > 0$.

Aplicando el Teorema de Bayes, la distribución conjunta a posteriori de β es

1. Bajo la distribución a priori normal plana:

$$p(\beta|y) \propto \exp \left\{ -\frac{1}{2}(\beta - \mathbf{b}_0)^t \mathbf{B}_0^{-1} (\beta - \mathbf{b}_0) + \sum_{i=1}^n y_i x_i^t \beta - \sum_{i=1}^n \log(1 + e^{x_i^t \beta}) \right\} \quad (3)$$

2. Bajo la distribución a priori no informativa:

$$p(\beta|y) \propto \exp \left\{ \sum_{i=1}^n y_i x_i^t \beta - \sum_{i=1}^n \log(1 + e^{x_i^t \beta}) \right\} \quad (4)$$

Note que las distribuciones a posteriori obtenidas no tienen una forma analítica conocida, por consiguiente los estimadores bayesianos de los parámetros no se

pueden obtener de forma directa. Tal situación propicia a utilizar alguna de las técnicas de Cadenas de Markov Monte Carlo. Gilks y Wild (1992) desarrollaron un algoritmo que permite muestrear de cualquier distribución absolutamente continua, con función de densidad log-cóncava, y que ha sido aplicado exitosamente en situaciones como la descrita antes, ver por ejemplo Dellaportas y Smith (1993). Para aplicar el algoritmo de Metropolis hay que seleccionar primero la distribución generadora de candidatos, en este paso hay que ser muy cuidadosos, ya que una elección inadecuada de dicha distribución puede provocar que el algoritmo tenga un desempeño muy pobre debido a la alta tasa de rechazos. En este trabajo, se utiliza este algoritmo para obtener muestras de las distribuciones a posteriori (3) y (4) y se selecciona a la distribución normal multivariada como la generadora de candidatos. Los libros de Carlin y Louis (2000) y Gelman *et al.* (2004) tienen abundantes ejemplos en los que se utiliza este algoritmo.

3. Estudio de Simulación

En el presente trabajo, el estudio de simulación se realiza con la ayuda de rutinas hechas en el programa R. Los datos se generan de la siguiente forma. La matriz diseño, $\mathbf{X} = [1, x_1^t, x_2^t]$, se generó de orden 3×40 . Los valores de x_{1t} se simularon con una distribución uniforme $[0, 1]$ y x_{2t} se obtuvo como $x_{2t} = x_{1t} + cu$; $i = 1, 2, \dots, n$, donde u tiene distribución uniforme $[0, 1]$ y c toma un valor apropiado para obtener una correlación muestral aproximada de 0.95. El β verdadero es dado por $\beta = 2VP1$, donde $VP1$ es el vector propio asociado al máximo valor propio de $\mathbf{X}^t\mathbf{X}$. La variable respuesta fue generada como $Y_t = 1$ si $\pi_t > U$ y $Y_t = 0$ en otro caso; donde U es una variable con distribución uniforme en el intervalo $[0, 1]$ y π_t generada como en (1). Para cada conjunto de datos generados se midió su traslape con la ayuda de la función `noverlap` de R y se clasificó en las categorías G0, G1 y G2 que corresponden al número de traslapes que están en los intervalos $[0, 1]$, $[2, 5]$ y $[6, 9]$ respectivamente.

Empleando el algoritmo de Metropolis se obtienen muestras Monte Carlo tanto de (3) y (4). Para la función generadora de candidatos se utiliza la distribución normal trivariada $N_3(\beta_R, \hat{\Sigma}_R)$, donde β_R es el estimador ridge iterativo y la matriz de varianzas y covarianzas $\hat{\Sigma}_R$ se obtiene mediante bootstrap con $B = 500$ repeticiones. Para arrancar el algoritmo de Metropolis se utilizan los estimadores de Ridge. Se fijan los valores de los hiperparámetros de la distribución a priori normal como $\mathbf{b}_0 = (0, 0, 0)^t$ y $\mathbf{B}_0 = \hat{\Sigma}_R \times 10^{200}$.

Se generaron $S = 10000$ muestras Monte Carlo de tamaño $M = 1000$ de las distribuciones (3) y (4), en cada muestra de Metropolis se obtiene los estimadores bayesianos bajo las funciones de pérdida lineal absoluta y la cuadrática que corresponden a la mediana y la media de la muestra a posteriori. Los estimadores se

comparan en términos de su sesgo y error cuadrático medio, con los estimadores de Ridge iterativo, EMV y los bayesianos propuestos por Gelman *et al* (2008). Los resultados se presentan en los Cuadros 1 y 2.

Cuadro 1: Estudio de sesgo para β_0 , β_1 y β_2 con $S = 10000$

| | | MV | RI | BNP1 | BNP2 | BU1 | BU2 | BC | BT | BNI |
|-----------|----|----------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| β_0 | G0 | ∞ | 0.529 | 1.107 | 0.978 | -7.391 | -7.651 | 0.801 | 0.768 | 0.753 |
| | G1 | 0.033 | -0.094 | 0.238 | 0.148 | -3.419 | -6.703 | -6.837 | -0.323 | -0.316 |
| | G2 | -0.683 | -0.722 | -0.582 | -0.622 | -0.151 | -0.090 | -0.704 | -0.705 | -0.705 |
| β_1 | G0 | ∞ | 0.534 | 1.774 | 1.834 | -1.605 | -1.370 | 0.068 | 0.007 | -0.014 |
| | G1 | 0.240 | 0.106 | 0.641 | 0.621 | -0.316 | -3.472 | 0.000 | -0.009 | -0.011 |
| | G2 | 0.176 | -0.047 | 0.464 | 0.385 | -1.042 | -0.596 | -0.101 | -0.080 | -0.075 |
| β_2 | G0 | ∞ | 0.558 | 1.944 | 1.853 | -5.116 | -5.260 | -0.228 | -0.285 | -0.302 |
| | G1 | 0.282 | 0.012 | 0.580 | 0.532 | -5.219 | -5.368 | -0.021 | -0.021 | -0.022 |
| | G2 | -0.557 | -0.407 | -0.563 | -0.537 | -7.119 | -7.348 | -0.504 | -0.492 | -0.490 |

Cuadro 2: Estudio de ECM para β_0 , β_1 y β_2 con $S = 10000$

| | | MV | RI | BNP1 | BNP2 | BU1 | BU2 | BC | BT | BNI |
|-----------|----|----------|-------|---------|---------|----------|----------|-------|-------|-------|
| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| β_0 | G0 | ∞ | 2.288 | 12.073 | 11.737 | 1243.304 | 1351.466 | 2.418 | 2.178 | 2.096 |
| | G1 | 3.371 | 1.075 | 4.041 | 3.737 | 728.436 | 799.513 | 0.939 | 0.941 | 0.940 |
| | G2 | 1.284 | 1.068 | 1.526 | 1.501 | 470.893 | 516.300 | 0.985 | 1.010 | 1.015 |
| β_1 | G0 | ∞ | 5.285 | 114.773 | 121.156 | 1692.763 | 1842.418 | 2.691 | 2.235 | 2.167 |
| | G1 | 70.998 | 3.991 | 74.871 | 74.875 | 1605.925 | 1760.713 | 1.875 | 1.966 | 1.994 |
| | G2 | 16.185 | 3.141 | 25.049 | 23.405 | 1655.240 | 1802.688 | 1.398 | 1.710 | 1.784 |
| β_2 | G0 | ∞ | 4.962 | 108.725 | 112.415 | 1588.068 | 1705.709 | 2.470 | 2.048 | 1.999 |
| | G1 | 62.108 | 3.453 | 64.327 | 63.860 | 1444.930 | 1574.200 | 1.714 | 1.805 | 1.830 |
| | G2 | 13.822 | 2.753 | 21.404 | 20.166 | 1425.558 | 1573.345 | 1.353 | 1.630 | 1.696 |

4. Conclusiones

El algoritmo de Metropolis resultó ser una herramienta adecuada para la estimación de parámetros en el caso de regresión logística cuando hay colinealidad en las variables explicativas y separación en los datos. Este algoritmo presenta varias ventajas sobre los métodos de estimación convencionales, ya que produce buenos resultados en casos en los que el método de máxima verosimilitud falla, no se requiere maximizar ninguna función, además de su fácil implementación.

En relación a los resultados vistos en este trabajo, se puede concluir que los estimadores bayesianos propuestos obtenidos a partir de la distribución a priori

conjunta normal plana, resultan ser competentes con relación a los propuestos por Gelman *et al* (2008). En el caso de las estimaciones calculadas bajo el enfoque clásico, el EMV y el estimador ridge iterativo se comportan de manera similar a los estimadores bayesianos, excepto el EMV en el grupo con menor traslape (G0) y en términos de error cuadrático medio, lo cual confirma la superioridad de los estimadores ridge iterativo y bayesianos sobre el EMV.

Referencias

- [1] Carlin, B. P. y Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman - Hall/CRC, New York, segunda edición.
- [2] Dellaportas P. y Smith A. F. M. (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling. *Applied Statistics*, Vol. 42, No. 3., pp. 443-459.
- [3] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80, 27-38.
- [4] Gelman, A.; Jakulin, A.; Grazia-Pittau M. y Yu-Sung Su (2008). A Weakly Informative Default Prior Distribution For Logistic And Other Regression Models. *The Annals of Applied Statistics* 2 (4), pp. 1360-1383.
- [5] Gelman, A., Carlin, J. B., Stern, H. S. y Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman - Hall/CRC, New York, segunda edición.
- [6] Gilks, W. R. y Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics*, Vol. 41, No. 2, pp. 337-348.
- [7] Heinze, G. y Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine* 21, 2409-2419.
- [8] Le Cessie, S. and J.C. van Houwelingen. 1992. Ridge estimators in logistic regression. *Applied Statistics*. 41(1):191-201.
- [9] Lee, A. H., and M. J. Silvapulle. 1988. Ridge estimation in logistic regression. *Communications in Statistics-Theory and Methods*. 17(4): 1231-1257.
- [10] R Development Core Team (2007). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- [11] Rousseeuw, P. J., and A. Christmann. 2003, Robustness against separation and outliers in logistic regression. *Computational Statistics and Data Analysis* 43, 315-332.
- [12] Schaefer, R. L., L. D. Roi, and R. A. Wolfe, 1984. A ridge logistic estimator. *Communications in Statistics-Theory and Methods*. 13(1): 99-113.

Un Estudio Comparativo de Estimadores de Índices de Diversidad: Una Corrección al Índice de Fager

C. Bouza¹, Di Rienzo, D. Covarrubias², L. Pimentel³

¹ Universidad de la Habana. Cuba. bouza@matcom.uh.cu

² Universidad Autónoma de Guerrero. México. dcova@uagro.mx

³ Maestría en Estadística. UAGro.

Resumen

La necesidad de determinar el grado de diversidad que guarda una comunidad ecológica ha sido un tema recurrente en ecología, esto ha llevado a la búsqueda de mecanismos que permitan cuantificarla ver Magurran (2004). Para realizar esta tarea la comunidad académica ha hecho uso de una herramienta matemática que mide la riqueza de especies y el equilibrio que guardan, los índices de diversidad. Kempton (2002) sugiere que la diversidad debe verse como parámetro poblacional el cual debe ser estimado. Se presentan estimadores de los índices de diversidad más utilizados por los ecólogos y biólogos: Simpson, Shannon y Fager. Se presenta una propuesta para corregir el índice de Fager y se realiza un estudio comparativo. El comportamiento de los estimadores propuestos es estudiado utilizando los datos de una investigación sobre Ordenamiento Ecológico llevado a cabo en Estado de Guerrero, México. Experimentos de Monte Carlo son desarrollados.

Palabras clave: *Biodiversidad, insesgidez, muestreo aleatorio simple, experimentos de*

Monte Carlo.

1. Introducción

El concepto de diversidad ha sido ampliamente discutido como se puede ver en la literatura especializada, autores como Magurran (2004), Krebs (1999), Gove et al (1994), entre otros coinciden al señalar que un índice de diversidad está formado por una función del número de especies y la abundancia. Un aporte a esta discusión la hacen Patil y Taille (1982) quienes definen a la diversidad como una propiedad medible e intrínseca de la comunidad ecológica y de acuerdo con Kempton (2002) esto permite considerar a la característica de interés como parámetro el cual debe ser estimado. Este es el objeto de estudio del presente trabajo.

Patil y Taille (1982) proponen caracterizar a una comunidad ecológica como el par $C(k, \pi)$ donde k es el número de especies conocidas que están físicamente presentes, y $\underline{\pi} = \{\pi_1, \pi_2, \dots, \pi_k\}$ es el vector de abundancia de especies. Se dice que

una comunidad está completamente equilibrada si $\pi_1 = \pi_2 = \dots = \pi_k = \frac{1}{k}$. En general en una población con N individuos $N = \sum_{i=1}^k N_i$ y $\pi_i = \frac{N_i}{N}$. Entonces es posible asociar $g(i, \pi)$ una medida de la rareza de la especie i , que dependería solamente del valor de π_i ya que quedaría sin cambio si otras especies son agrupadas dentro de una categoría complementaria.

Definición 1 Patil y Taille (1989) El índice de diversidad $\lambda(C)$ asociado a una comunidad $C(k, \underline{\pi})$ es definido como el promedio de la rareza

$$\lambda(C) = \sum_{i=1}^k \pi_i g(i; \underline{\pi}) \quad (1)$$

donde $\underline{\pi}$ es el vector de abundancias de las especies físicamente presentes y el valor máximo del índice será cuando la comunidad esté en equilibrio

Patil y Taille (1982) proponen clasificar los índices de diversidad en dicotómicos y de rango, por otra parte Bouza y Covarrubias (2005) proponen dos clases generales a partir de su estructura: la *clase aditiva* \wedge^+ y la *clase multiplicativa* \wedge^\times esto permite hacer un tratamiento unificado del problema de la estimación.

La clase aditiva la definida como:

$$\wedge^+ = \left\{ \lambda = \sum_{i=1}^k (\tau_i g(i, \pi_i)) \pi_i \right. \\ \left. : \pi_i = \frac{N_i}{N}, N > 0 \right\} \quad (2)$$

y la clase multiplicativa por:

$$\wedge^\times = \left\{ \lambda = \prod_{i=1}^k (\tau_i g(i, \pi_i)) \pi_i \right. \\ \left. : \pi_i = \frac{N_i}{N}, N > 0 \right\} \quad (3)$$

donde: τ_i se asocia el rango que se le asigna a la especie i , $g(i, \pi_i)$ la medida de rareza de especie y π_i la abundancia relativa de la especie i .

Este trabajo se encuentra dividido en cuatro secciones, la segunda describe el índices de diversidad de Fager para un sitio de muestreo, esto permite desarrollar las ideas. La tercera presenta el estimador del índice de Fager transformado para un sitio de muestreo. Por último el se muestran los resultados de los experimentos de Monte Carlo desarrollados con los datos obtenidos por el Centro de Investigación Área de Ciencias Naturales de la Universidad Autónoma de Guerrero, México.

2. Índice de Fager

Fager (1972) propuso un índice llamado "número de transferencias", está basado en el número de transferencias que se deben hacer para que la distribución de individuos sea equilibrada. Para ello es necesario la asignación de rangos de la abundancia de especie. De acuerdo con la clasificación de propuesta por Patil y Taille (1982) este índice pertenece a los índices de rango. El índice de Fager está definido por:

$$\lambda_{Fager}^* = \frac{N(k+1)}{2} - \sum_{i=1}^k N_i R_i \quad (4)$$

donde: k es el número de especies, N_i el número de individuos de la especie i en la comunidad, N el número total de individuos. R_1, \dots, R_k son los rangos de las especies en orden decreciente de su importancia, es decir que el rango para la especie de mayor importancia se le asigna rango uno.

Note que se presenta un problema cuando se desea obtener un estimador de (4) ya que se tiene que estimar N , por ello Bouza y Schubert (2002) proponen una transformación del índice de Fager para estimar las abundancias relativas π_i con $\hat{\pi}_i = \frac{n_i}{n}$ donde n_i es el número de individuos de la especie i en los n observados $n = \sum_{i=1}^k n_i$ realizando la transformación es fácil observar que al lograr el equilibrio entre las especies de una comunidad, es decir $\pi_i = \frac{1}{k}$, el valor

que obtenemos para el índice es cero, por lo que la interpretación como índice de diversidad se debe de hacer en forma inversa a los índices más utilizados como Shannon y Simpson. Al describir el índice (4) es posible ubicarlo en la clase aditiva (2).

3. Transformación del índice de Fager

Gove et al (1994) refieren como un problema el tener un rango diferente para cada índice, lo que hace menos sencilla la tarea de los especialistas para la comparación de los valores obtenidos. Una propuesta para transformar el índice de Fager es

$$\lambda_{FagTrans} = \frac{1}{\lambda_0} \sum_{i=1}^k \pi_i R_i \quad (5)$$

donde: k es el número de especies, $\lambda_0 = \frac{k+1}{2}$, $\pi_i = \frac{N_i}{N}$ y R_1, \dots, R_k son los rangos de las especies en orden decreciente de su importancia.

De esta manera el índice se encuentra acotado en el intervalo $(0, 1]$, el máximo valor lo obtiene cuando las especies se encuentran en equilibrio y su interpretación es similar a la de los índices mas utilizados.

Supondremos que se selecciona un sitio de muestreo y las variables aleatorias son producidas por un procedimiento equivalente al *Muestreo Aleatorio Simple [MAS]*. Note que cuando se usa un muestreo aleatorio simple *con reemplazo [MAScr]* para n fija tenemos que las $n_i \sim B(n, \pi_i)$ donde las n_i son los elementos de la especie i que se tiene de la muestra y π_i es la probabilidad de encontrar un individuo de la especie i en la muestra.

Proposición 2 *Un estimador insesgado del índice transformado de Fager (5), su varianza y el estimador de la varianza estimador en el MAScr,*

suponiendo n fija están dados por:

$$\hat{\lambda}_{FT} = \frac{1}{\lambda_0} \sum_{i=1}^k \hat{\pi}_i R_i \quad (6)$$

$$Var\{\hat{\lambda}_{FT}\} = \frac{1}{n^2 \lambda_0^2} \left[\sum_{i=1}^k \pi_i R_i^2 - \left(\sum_{i=1}^k R_i \pi_i \right)^2 \right] \quad (7)$$

$$\widehat{Var}\{\hat{\lambda}_{FT}\} = \frac{1}{n^2 \lambda_0^2} \left[\sum_{i=1}^k R_i^2 \hat{\pi}_i + \sum_{i=1}^k \frac{R_i^2 \hat{\pi}_i (1 - n \hat{\pi}_i)}{n-1} - \frac{2n}{n-1} \sum_{i < j} R_i R_j \hat{\pi}_i \hat{\pi}_j \right] \quad (8)$$

donde $\lambda_0 = \frac{k+1}{2}$, los R_i son rangos asignados a las especies y $\hat{\pi}_i = \frac{n_i}{n}$.

4. Simulación

Para realizar la simulación se utilizó la base de datos obtenida en la investigación Ordenamiento Ecológico Territorial desarrollada en la subcuenca del río San Juan al norte del Estado de Guerrero. Como una breve descripción de la subcuenca se puede apuntar que el recorrido de la subcuenca del río de aproximadamente 46 kms tocando cinco Municipios del Estado, ver Almazán et al (2003). La muestra para el estudio de vegetación la conforma 60 transectos asignados proporcionalmente al tamaño de la superficie que ocupa cada tipo de vegetación identificado. Se colectaron 875 ejemplares e identificaron 400 géneros comprendidas en 103 familias.

Los valores de los índices de diversidad, sus varianzas de Simpson, Shannon y Fager son calculados utilizando los resultados de Bouza y Covarrubias (2005) y Cruz (2009). Note que la estimación del índice transformado de Fager tiene un mejor comportamiento que el índice original ver cuadro (1), lo que sugiere que tiene un comportamiento congruente con lo que expresan los otros valores y en consecuencia puede ser utilizado como una medida que aproxima el valor de la diversidad de una comunidad ecológica.

| Índice | Estimadores | | | | error | [9] Magurran A. E. (2004), "Measuring Biological Diversity," Blackwell Publishing. |
|--------------|-------------|-----------------|-----------|---------------------------|--------|---|
| | λ | $\hat{\lambda}$ | \hat{V} | $\{\hat{\lambda}_{F_x}\}$ | | |
| Simpson | 0.9583 | 0.9556 | 0.0002 | 0.0027 | 0.0027 | [10] Patil G. P. and Taille C. (1982), "Diversity as a Concept and its Measurement," Journal of the American Statistical Association 77, 548-567. |
| Shannon | 3.6953 | 3.4526 | 0.0023 | 0.2427 | 0.0703 | |
| Fager | 16.1929 | 7.8422 | 2.9367 | 8.3508 | 1.0640 | |
| Fager transf | 0.5166 | 0.5624 | 0.2075 | 0.0458 | 0.0814 | |

Cuadro 1: Estimaciones de los índices de diversidad

Referencias

- [1] Almazán A., Urbán G., González R., Tapia J., Villerías S., Beltrán E., Almazán M. (2003), "Ordenamiento Ecológico Territorial de la Subcuenca del Río San Juan del Estado de Guerrero," Reporte Técnico de Investigación, SIBEL-UAGro.
- [2] Bouza C. y Covarrubias D. (2005), "Estimación del índices de diversidad de Simpson en m sitios de muestreo," Investigación Operacional, Vol. 26, No.2, pp 187-195.
- [3] Bouza C. y Schubert L. (2002), "The estimation of Bio-diversity and the characterization of the dynamics: an application to the study pest," Rev. Matematica e Estadística. 21, 85-98.
- [4] Cruz U. (2009), "Estimaciones del índice de bio-diversidad de Shannon" Maestría en estadística aplicada UAGro Tesis.
- [5] Fager E.W. (1972), "Diversity: A Sampling Study," American Naturalist, 106, 293-310.
- [6] Gove J. H., Patil G.P., Swindel B.F. and Taille C. (1994), "Ecological Diversity and Forest Management," Handbook of Statistical, Vol 12, (eds G.P. Patil and C.R. Rao, Elsevier Science B.V.)
- [7] Kempton, R. (2002), "Species diversity", Encyclopedia of Environmetrics, 4, 2086-2092. (eds Abbel H. El-Shaarawi and W. Piegorsh) John Wiley & Sons. N. York
- [8] Krebs C. (1999), "Ecological Methology," 2nd Edition, Adison Wesley, London.

Comparación de la Arquitectura de dos Especies Simpátricas de Cactáceas Columnares *Cephalocereus palmeri* var. *sartorianus* y *Neobuxbaumia scoparia* en el Bosque Seco Caducifolio de Cerro Gordo, Veracruz

Judith Guadalupe Montero Mora¹, Armando J. Martínez², Sergio Francisco Juárez Cerrillo³

¹Fac. de Estadística e Informática. Universidad Veracruzana. Xalapa, Ver. jmontero@uv.mx

²Instituto de Neuroetología. Universidad Veracruzana, Xalapa, Ver. armartinez@uv.mx

³Fac. de Estadística e Informática. Universidad Veracruzana. Xalapa, Ver sejuarez@uv.mx

Resumen

Sean X y Y dos dimensiones que cambian con el tamaño en las especies 1 y 2. Supongamos X y Y satisfacen las relaciones $Y = f(X)$ y $Y = g(X)$ en las especies 1 y 2, respectivamente. Si X y Y cambian de manera similar en ambas especies, entonces tenemos que $f = g$. En este trabajo presentamos una prueba para probar la hipótesis $H_0 : f = g$ vs. $H_a : f \neq g$ y la aplicamos para comparar la alometría de dos especies de cactáceas del bosque seco caducifolio de Cerro Gordo, en el estado de Veracruz. Los datos analizados consisten de la altura y el perímetro de la planta. El procedimiento que utilizamos es no-paramétrico en el sentido de que no requiere que se establezca la forma de las funciones f y g . Además la prueba es consistente y está basada en la discrepancia entre f y g medida en términos de la distancia ponderada L^2 .

Palabras Clave: Alometría, regresión no paramétrica, cactáceas.

1. Introducción

Las plantas habitan ambientes diversos y se caracterizan por tener una amplia variedad de adaptaciones morfológicas, arquitectónicas y fisiológicas. En particular algunas suculentas habitan ambientes donde la disponibilidad de agua es temporalmente intermitente, quedando así sujetas a períodos variables de desecación. Tal es el caso de las cactáceas y árboles que habitan en selvas bajas caducifolias, las cuales cumplen un importante papel dentro de las comunidades vegetales, ya que interactúan con múltiples factores bióticos y abióticos que influyen sobre el crecimiento y desarrollo de las plantas (Bullock, 2000). Es así que la aproximación alométrica intra e interespecífica permite evaluar parte de las respuestas de las plantas a la exposición a presiones selectivas impuestas por el ambiente (Niklas 1994, King *et al.*, 2006). Por lo que, en este estudio presentamos un análisis alométrico entre dos especies de cactáceas columnares que cohabitan en un fragmento de bosque seco caducifolio: *Cephalocereus palmeri* var. *sartorianus* y *Neobuxbaumia scoparia*, a partir de la relación entre la altura y diámetro de la planta.

Muchos autores han analizado las relaciones alométricas utilizando modelos de regresión lineal y no lineal, para pares de variables y con análisis de covarianza cuando desean hacer

comparaciones de las pendientes de la regresión para comparar entre especie o poblaciones (Niklas, 1994; Thomas 1996). El objetivo de este trabajo es proponer un procedimiento de prueba de hipótesis para comparar las relaciones alométricas basado en las diferencias de dos funciones de regresión. Las hipótesis que se plantean son $H_0 : f = g$ vs. $H_a : f \neq g$, es decir se pretende comparar dos curvas de regresión correspondientes a cada una de las especies bajo estudio, con tamaños de muestra m, n ($m \leq n$). El procedimiento utilizado es no-paramétrico ya que no se establece paraméricamente la forma de las funciones f y g .

2. Método

Los datos que se analizan fueron colectados en un fragmento de bosque seco caducifolio de Cerro Gordo, Veracruz. El interés se centra en evaluar la altura de las cactáceas en relación con su diámetro, comparando por especie. Los datos consisten entonces de dos muestras independientes: *C. palmeri* var. *sartorianus* $m = 77$ y *N. scoparia* $n = 110$.

La comparación alométrica de estas dos especies la realizamos probando las siguientes hipótesis

$$H_0 : f = g \text{ vs. } H_a : f \neq g \quad (1)$$

donde $Y = f(X)$ y $Y = g(X)$ son las relaciones que satisfacen las dimensiones X y Y de las especies 1 y 2, respectivamente. De modo que si X y Y cambian de manera similar en ambas especies, entonces tenemos que $f = g$. Si f y g fueran lineales, entonces la opción es un análisis de covarianza. Sin embargo, esto tiene restricciones ya que las suposiciones de un modelo lineal o la suposición de normalidad pueden no ser realistas. Por lo que proponemos aplicar el procedimiento de prueba propuesto Munk y Dette (1998) para probar la hipótesis en (1).

Se consideraron en el análisis las dos muestras independientes de $m = 77$ y $n = 110$ observaciones por lo que supondremos que

$$Y(X_{1,i}), Y(X_{2,j}) \in L^2, \quad i = 1, \dots, 77, \quad j = 1, \dots, 110 \quad (2)$$

Sin pérdida de generalidad, supondremos que el rango de la variable regresora X es el intervalo $[0,1]$. Consideramos los siguientes modelos

$$\begin{aligned} Y_i = Y(X_{1,i}) &= f(X_{1,i}) + \varepsilon(X_{1,i}) = f_i + \varepsilon_i, \quad i = 1, \dots, 77, \\ Y_j = Y(X_{2,j}) &= g(X_{2,j}) + \eta(X_{2,j}) = g_j + \eta_j, \quad j = 1, \dots, 110 \end{aligned} \quad (3)$$

Supondremos que las funciones de regresión y las variables aleatorias son tales que $f, g \in L^2[0,1]$ y $\varepsilon, \eta \in L^2$. Supondremos también que

$$E[\varepsilon(X)] = E[\eta(X)] = 0, \quad V[\varepsilon(X)] = \sigma_\varepsilon^2(X), \quad V[\eta(X)] = \sigma_\eta^2(X) \quad (4)$$

El procedimiento de comparación consiste en estimar una medida de discrepancia entre las dos funciones f y g usando la distancia L^2

$$D^2 = \|f - g\|^2 = \int_0^1 [f(x) - g(x)]^2 dx \quad (5)$$

Por lo que la hipótesis en (1) es equivalente a:

$$H_0 : D^2 = 0 \text{ vs } H_a : D^2 \neq 0 \quad (6)$$

En el caso de diferentes tamaños de muestra, la propuesta para una estimación de D^2 es la siguiente:

$$\hat{D}^2 = \sum_{i=0}^m \sum_{j=0}^n \lambda_{ij} (Y_{i+1} - Y_{j+1})(Y_i - Y_j) \quad (7)$$

donde $Y_{10} = Y_1$, $Y_{20} = Y_1$ y $Y_{m+1} = Y_m$, $Y_{n+1} = Y_n$, $m, n \geq 1$. Los pesos λ_{ij} están dados por:

$$\begin{aligned} \lambda_{ij} &= \int_0^1 \mathbf{I}_{[X_{1,i}, X_{1,i+1})}(X) \mathbf{I}_{[X_{2,j}, X_{2,j+1})}(X) dX \\ &= \min(X_{1,i+1}, X_{2,j+1}) - \max(X_{1,i}, X_{2,j}) \mathbf{I}_{\{S\}}(X) \end{aligned} \quad (8)$$

con

$$\mathbf{I}_{\{S\}}(X) = \begin{cases} 1 & \text{si } X \in S = \{\min(X_{1,i+1}, X_{2,j+1}) > \max(X_{1,i}, X_{2,j})\} \\ 0 & \text{si } X \notin S \end{cases}$$

Una vez calculado el valor \hat{D}^2 , el estadístico de prueba propuesto es el siguiente:

$$T_{m,n} = \sqrt{n+m} \hat{D}^2 \sim N(0, \delta^2) \quad (9)$$

donde $N(0, \delta^2)$ denota la distribución normal centrada con varianza δ^2 dada por:

$$\delta^2(D^2) = \frac{\sigma_\varepsilon^4}{k} + \frac{\sigma_\eta^4}{k} + 2\sigma_\varepsilon^2 \sigma_\eta^2 \ell(r, s) \quad (10)$$

Con $k = m/(n+m)$. El cociente de los tamaños de muestra muestras puede reescribirse como $m/n = r/s \leq 1$, tal que el máximo común divisor de r y s $mcd(r,s) = 1$, si esto se cumple, se tiene que para enteros $n \geq m$

$$\ell(r,s) = (r+s) \frac{1-r^2+3rs}{3rs^2} \quad (11)$$

La estimación de las varianzas σ_ε^2 y σ_η^2 propuestas en el caso de regresión no paramétrica es:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2(m-1)} \sum_{i=2}^m (Y_i - Y_{i-1})^2, \quad \hat{\sigma}_\eta^2 = \frac{1}{2(n-1)} \sum_{j=2}^n (Y_j - Y_{j-1})^2$$

3. Resultados

La relación entre el diámetro y la altura varió entre las dos especies de cactáceas (Fig. 1). Es evidente que la especie *C. palmeri* var. *sartorianus*, muestra mayor altura comparado a la respuesta que tiene *N. scoparia*

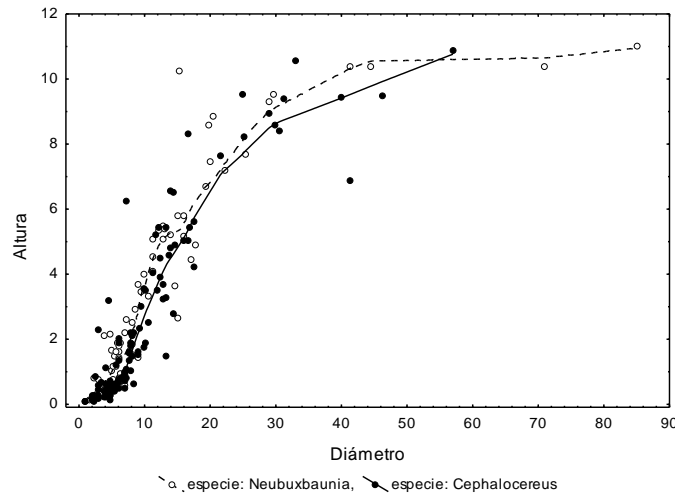


Figura 1. Relación entre altura vs. diámetro de las cactáceas columnares *Neubuxbaunia scoparia* y *Cephalocereus palmeri* var. *sartorianus*.

Para el cálculo del estadístico de prueba se cuenta con $n=110$ de observaciones *N. scoparia* (especie 1) y $m=77$ de la especie *C. palmeri* var. *sartorianus*, por lo que $(r,s) = (77,110)$, lo que dio por resultado $\ell(r,s) = 1.3034$. Las varianzas de la altura de cada especie son $\hat{\sigma}_\varepsilon^2 = 0.7029$ y de $\hat{\sigma}_\eta^2 = 1.036$. El valor de $\hat{\delta}^2 = 4.925$ y el valor de la

distancia es $\hat{D}^2 = 2.784$ por lo que $T_{m,n} = \sqrt{n+m}\hat{D}^2 / \hat{\delta}^2 = 17.155$, este valor indica que hay muy fuerte evidencia de la diferencia entre las alturas de las dos especies.

4. Conclusión

La relación alométrica entre el diámetro y altura de las cactáceas columnares indica claras diferencias entre especies. El estadístico de prueba propuesto llevó al rechazo de la hipótesis nula lo que confirma que el tamaño de las cactáceas difiere interespecíficamente.

REFERENCIAS

Bullock, S. (2000). Developmental patterns of tree dimensions in a neotropical deciduous forest. *Biotropica* 32: 42-52.

King, D. Wright S., Connell J. (2006). The contribution of interspecific variation in maximum tree height to tropical and temperate diversity. *Journal of Tropical Ecology* 22: 11-24.

Munk, A., Dette, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory. *The Annals of Statistics* 26: 2339-2368.

Niklas K. (1994). *Plant allometry: the scaling of form and process*. University of Chicago. 395 pp.

Thomas S. (1996). Asymptotic height as a predictor of growth and allometric characteristics in Malaysian rain forest trees. *American Journal of Botany* 83: 556-566.

Un Modelo de Regresión Poisson Inflado con Ceros para Analizar Datos de un Experimento de Fungicidas en Jitomate

María de Lourdes Velasco Vázquez¹, Sergio Francisco Juárez Cerrillo²

¹ Universidad Veracruzana. México. mlvv10@hotmail.com

² Universidad Veracruzana. México. sejuarez@uv.mx

Introducción

Las variables respuesta que son conteos son frecuentemente medidas en las ciencias biológicas. Sin embargo, no es raro que el número de ceros observados en la variable respuesta exceda a la frecuencia que se espera observar bajo la distribución que se ajusta. En este caso se indica que los datos presentan exceso de ceros o están inflados con ceros. De ahí que está alta frecuencia en muchos casos sea especificada incorrectamente al modelo estadístico, ya que no se cumplen los supuestos distribucionales de la modelación. Lo cual, no debe ser ignorado ni tampoco debe ajustarse una distribución con esta característica ya que las inferencias y niveles de significancia y confianza no serán correctos.

El exceso de ceros es común en la práctica, por lo que se han desarrollado modelos estadísticos que describen este fenómeno y, por lo tanto, permiten derivar conclusiones realistas y confiables a partir de las inferencias. Este sentido, la aplicación de este tipo de modelación no se ha extendido tan ampliamente en las ciencias biológicas. Aunque su uso esta asociado a los estudios de abundancia y presencia de especies (Barry y Welsh 2002;), reclutamiento de árboles (Fortin y DeBlois 2007).

Es así como en este estudio se analizaron datos obtenidos en un experimento de pruebas de fungicidas para el control de daños en plantas de jitomate. El experimento se realizó en condiciones de invernadero bajo un diseño anidado con tres tratamientos (fungicidas) y un testigo. Las variables respuesta fueron el número de daños que presentaron las plantas de jitomate en foliolos, hojas, racimos y frutos. El problema de investigación queda planteado en dos puntos: determinar si los fungicidas significativamente reducen el número de daños. En caso afirmativo, determinar cual fungicida es el más efectivo. Debido a que los datos presentaban exceso de ceros, se utilizó el modelo de regresión Poisson inflado con ceros de Lambert para cada variable explicatoria, ajustados con el programa STATA.

En la Figura 1 se muestra la distribución de frecuencias del número total de daños en hojas. Destaca inmediatamente la alta frecuencia de plantas que no presentaron daño, lo cual indica exceso de ceros. Este comportamiento prevalece en las variables respuesta número de daños en foliolos, racimo y fruto.

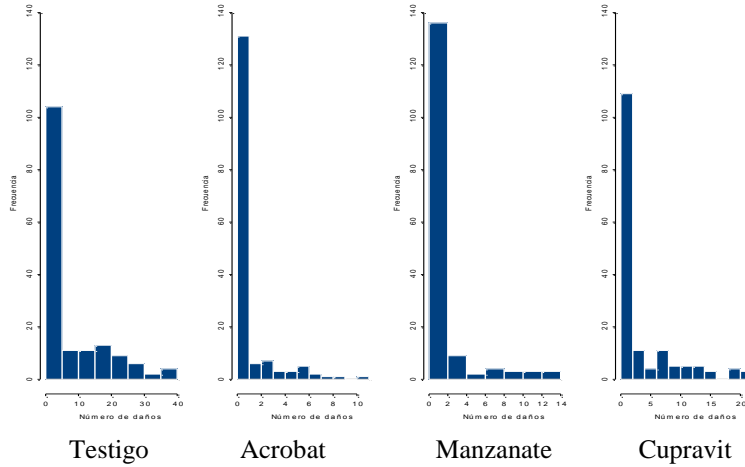


Figura 1. Distribución del número total de daños en Foliolos y Hojas.

El Modelo de Regresión PIC

En los modelos PIC se mezcla una distribución degenerada en el cero con una distribución discreta no degenerada. La estructura de la regresión se construye a través de la media de la distribución no degenerada y, posiblemente, a través de mezclar las probabilidades. A continuación presentamos el modelo de regresión Poisson inflado con ceros desarrollado en Lambert (1992).

Sea $Y = (Y_1, \dots, Y_n)^T$ el vector de la variable respuesta. El modelo de regresión PIC supone que las Y_i son independientes con la siguiente distribución.

$$Y_i = 0 \text{ con probabilidad } p_i,$$

$$Y_i \sim \text{Poisson}(\lambda_i) \text{ con probabilidad } 1 - p_i.$$

De tal manera que la función de masa de probabilidad de Y_i es

$$P(Y_i = y) = \begin{cases} p_i + (1 - p_i)e^{-\lambda_i}, & \text{para } y = 0, \\ (1 - p_i) \frac{e^{-\lambda_i} \lambda_i^y}{y!}, & \text{para } y = 1, 2, \dots \end{cases}$$

Denotamos lo anterior por $Y_i \sim \text{PIC}(p_i, \lambda_i)$. La variable respuesta se modela mediante los parámetros del modelo $\lambda = (\lambda_1, \dots, \lambda_n)^T$ y $\mathbf{p} = (p_1, \dots, p_n)^T$. Se supone que estos parámetros dependen de las covariables a través de un modelo lineal generalizado. Es decir, se supone que

$$\log(\lambda) = (\log(\lambda_1), \dots, \log(\lambda_n))^T = \mathbf{B}\beta,$$

$$\text{logit}(\mathbf{p}) = (\text{logit}(p_1), \dots, \text{logit}(p_n))^T = \mathbf{G}\gamma,$$

donde β y γ son parámetros desconocidos, y \mathbf{B} y \mathbf{G} son matrices con las covariables. Cuando los parámetros γ y β no están relacionados y las covariables que afectan a la distribución Poisson son las mismas covariables que afectan a las probabilidades p_i , la log-verosimilitud para el modelo de regresión PIC es

$$\begin{aligned} L(\gamma, \beta; Y) = & \sum_{i=1}^n \log p_i + \sum_{Y_i=1} \log\{\exp(\mathbf{G}_i^T \gamma) + \exp[-\exp(\mathbf{B}_i^T \beta)]\} \\ & + \sum_{Y_i>0} [Y_i \mathbf{B}_i^T \beta - \exp(\mathbf{B}_i^T \beta)] - \sum_{Y_i>0} \log(Y_i!), \end{aligned}$$

donde \mathbf{G}_i y \mathbf{B}_i son las filas i de \mathbf{G} y \mathbf{B} , respectivamente. Notemos que

$$p_i = \frac{\exp(\mathbf{G}_i^T \gamma)}{1 + \exp(\mathbf{G}_i^T \gamma)}$$

Así que la log-verosimilitud del modelo de regresión PIC es

$$\begin{aligned} L(\gamma, \beta; Y) = & \sum_{i=1}^n \mathbf{G}_i^T \gamma - \sum_{i=1}^n \log(1 + \exp(\mathbf{G}_i^T \gamma)) + \sum_{Y_i=0} \log\{\exp(-\mathbf{G}_i^T \gamma) + \exp[-\exp(\mathbf{B}_i^T \beta)]\} \\ & + \sum_{Y_i>0} [Y_i \mathbf{B}_i^T \beta - \exp(\mathbf{B}_i^T \beta)] - \sum_{Y_i>0} \log(Y_i!). \end{aligned}$$

los estimadores de máxima verosimilitud se obtuvieron a través del algoritmo EM.

Aplicación del modelo

Sea Y_{ijk} la variable aleatoria número total de daños en la variable respuesta i de la planta k bajo el tratamiento j ; donde $i = 1, \dots, 4$, con 1 daño en folíolos, 2 daño en hojas, 3 daño en racimo y 4 daño en fruto; $k = 1, \dots, 160$, pues ignoramos al cuadrante, lo que nos resulta en $5 \times 32 = 160$ plantas; $j = 1, \dots, 4$, con 1 acrobat (Ac), 2 manzanate (Mz), 3 cupravit (Cu), y 4 el testigo.

El modelo que ajustamos a los conteos es un modelo de regresión PIC. Suponemos que $Y_{ijk} \sim \text{PIC}(p_{ij}, \lambda_{ij})$, además que las observaciones entre plantas son independientes e idénticamente distribuidas ($k = 1, \dots, 160$). Los parámetros del modelo de regresión PIC son

$$\log(\lambda_{ij}) = \beta_{0ij} + \beta_{1ij} \text{Ac} + \beta_{2ij} \text{Mz} + \beta_{3ij} \text{Cu} ,$$

$$\text{logit}(p_{ij}) = \gamma_{0ij} + \gamma_{1ij} \text{Ac} + \gamma_{2ij} \text{Mz} + \gamma_{3ij} \text{Cu} ,$$

donde la variable explicatoria del modelo es la variable indicadora del fungicida. Las estimaciones de los parámetros de los modelos de regresión PIC se muestran en la Tabla 1.

| VARIABLES | Foliolo | Hoja | Racimo | Fruto |
|----------------|----------------------|---------------------|---------------------|---------------------|
| Poisson | | | | |
| Constante | 4.2365 (0.0131) | 2.2233 (0.0390) | 1.5359 (0.0637) | 1.5359 (0.0637) |
| Ac | -1.0108 (0.0337) | -0.8727 (0.0971) | -0.6004 (0.1766) | -1.3746 (0.2568) |
| Mz | -0.8733 (0.0294) | -0.7068 (0.0826) | -0.4480 (0.1278) | -0.8116 (0.1775) |
| Cu | -0.3281 (0.0212) | -0.0679 (0.0588) | 0.1973 (0.0863) | -0.1768 (0.1061) |
| Ceros | | | | |
| Constante | -0.10008 (0.1583) | 0.2257 (0.1591) | 0.6318 (0.1673) | 0.6318 (0.1673) |
| Ac | 1.1656 (0.2405) | 1.01984 (0.2498) | 1.4061 (0.3111) | 0.8787 (0.3444) |
| Mz | 0.8885 (0.2327) | 0.7607 (0.2395) | 0.7280 (0.2637) | 1.0450 (0.2945) |
| Cu | 0.3007 (0.2243) | 0.2847 (0.2280) | 0.0654 (0.2373) | 0.5074 (0.2511) |

Tabla 1. Ajuste del modelo PIC. Los valores entre paréntesis son los errores estándar.

En el modelo de regresión PIC todos los coeficientes son significativos en las cuatro variables respuesta, además los signos del modelo Poisson son negativos. En la parte de excesos de ceros, también los coeficientes resultaron significativos, pero con signo positivo. Lo que significa que la probabilidad del proceso que genera sólo los ceros aumenta conforme disminuye la media del proceso Poisson. Así, se establece que con los fungicidas Ac y Mz contienen mas ceros en las variables respuesta, ocasionando una disminución en el número de daños.

Para comparar el modelo de regresión Poisson con el modelo PIC se realiza una comparación de las distribuciones ajustadas. En la Figura 2 se muestra las distribuciones ajustadas para la variable daños en hojas. De esta inspección es claro que el modelo Poisson subestima el número de ceros y que el modelo PIC proporciona un mejor ajuste a los datos.

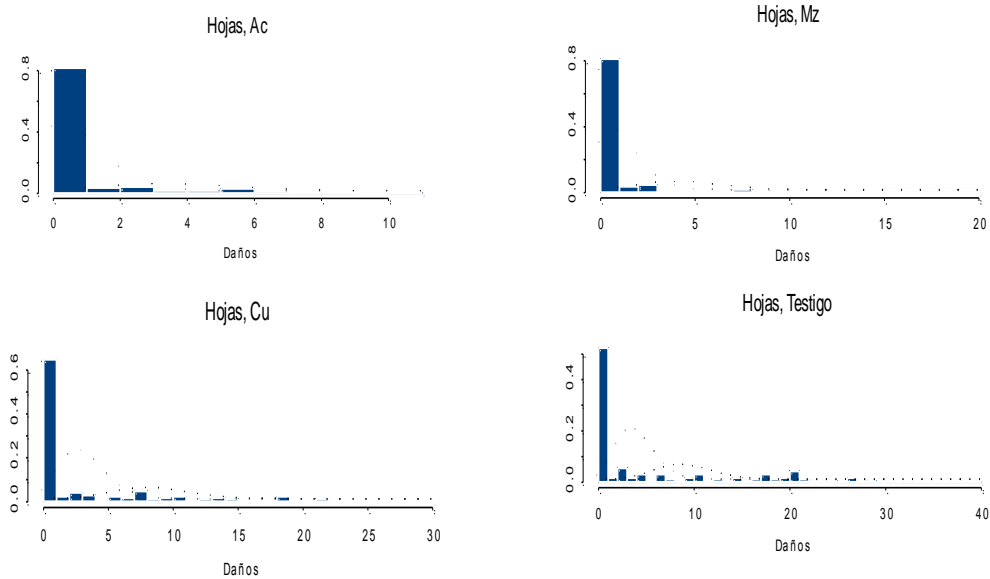


Figura 2. Distribuciones Poisson (círculos) y PIC (círculos rellenos) ajustadas y observadas del número de daños en hojas para cada fungicida.

Conclusiones

Los resultados obtenidos del análisis nos permiten concluir, en primer lugar, que el modelo de regresión Poisson no es la herramienta adecuada para los datos debido al exceso de ceros. Esto se refleja en el ajuste pobre que proporciona a los datos, en particular a los ceros. Por otro lado, el modelo PIC modela adecuadamente al exceso de ceros y proporciona en general un buen ajuste al número de daños en folíolos, hojas, racimos y frutos. Los modelos ajustados, ponen en evidencia una relación negativa y significativa entre el número de daños y el uso de fungicida. Esto es, el número de daños es menor cuando se usa fungicida que cuando no se usa. Más aún, el modelo de regresión PIC indica que los fungicidas que dan los mejores resultados para el control de daños son Ac y Cu en las cuatro variables respuesta.

Agradecimientos: Agradecemos al agrónomo Giovani Guadalupe Molina López que amablemente nos permitió usar sus datos de experimentación en jitomate en este trabajo.

Referencias

1. Hall, D.B., and Zhang, Z. (2004). Marginal models for zero inflated clustered data. *Statistical Modelling*, 4, 161-180.
2. Heilbron, D. C. (1994). Zero altered and other regression models for count data with added zeros. *Biometrics*, 36, 531-547.
3. Lambert, D. (1992). Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing. *Technometrics*, 34, 1-14.

4. Fortin, M and Deblois J. (2007). Modeling tree recruitment with zero-inflated model: the example of hardwood stands in Southern Québec, Canada. *Forest Science* 53, 529-539.P
5. Podlich, H.M., Faddy, M.J. & Smyth, G.K. (2002). A general approach to modeling and analysis of species abundance data with extra zeros. *Journal of Agriculture, Biology, and Environmental Statistics*, 7, 324-334.

Una Propuesta para Obtener la Estimación Máximo Verosímil cuando la Ecuación de Verosimilitud Presenta una Estructura Compleja

Ernesto Menéndez Acuña

Facultad de Matemáticas. Universidad Veracruzana. emenendez@uv.mx

1. Introducción

La estimación máximo verosímil casi siempre transita por la solución de la denominada ecuación de verosimilitud, (una ecuación o sistema de ecuaciones cuando se trate de la estimación de más de un parámetro). Suele ocurrir que dicha ecuación de verosimilitud presente una estructura compleja que no permita una solución analítica exacta. Esta situación se puede producir por las características del problema de estimación o por tratarse de un problema de datos incompletos. Cualquiera sea el caso, la solución de la ecuación de la verosimilitud exige el empleo de un método numérico.

Entre los métodos numéricos existentes, uno muy conocido es el de Newton o alguna de sus variantes, pero existen más que permiten resolver ecuaciones lineales o no lineales. Otro algoritmo que no deja de ser numérico, pero que su fundamento es más desde una perspectiva estadística, es el algoritmo EM (Dempster, Laird y Rubin, 1977) o algunas de sus extensiones. Para más información puede consultarse a McLachlan y Krishnan (1997) y Quintana, Villalobos y Cornejo (2005).

En este trabajo se presenta una alternativa para encontrar, cuando sea posible, la solución de la ecuación de verosimilitud sin necesidad de utilizar un método numérico. No se trata de negar la utilidad de éstos, sino de incluir una herramienta más para la solución de la ecuación de verosimilitud cuando ésta presente una estructura compleja y posibilite la aplicación del procedimiento que se propone.

El fundamento de este procedimiento consiste en evaluar numéricamente la función de verosimilitud o el logaritmo de esta función en un intervalo o región dentro del conjunto de posibles valores del parámetro y posteriormente determinar el valor máximo de la misma, así como el valor del parámetro (o parámetros) que produce este valor máximo, el cual por definición, es la estimación máximo verosímil. Sin el uso de las facilidades de cómputo actuales, esta propuesta no tiene sentido.

A la luz de lo que se ha podido revisar, este método se desempeña bien cuando solamente existe un parámetro objeto de estimación. En el caso de dos parámetros requiere de un poco más de esfuerzo computacional. En el caso de más de dos parámetros no se ha logrado una implementación eficiente.

Para ilustrar la propuesta se utilizan tres ejemplos que a su vez caracterizan tres situaciones en la estimación máximo verosímil de parámetros. El primero de ellos muestra una ecuación de verosimilitud que requiere en su solución de un método numérico. Es un caso

típico, quizás en extremo, de un problema de datos incompletos. Se trata de una muestra aleatoria de tamaño n , de la cual solamente se conoce su tamaño y el máximo de las observaciones. Se asume como modelo probabilístico el modelo Exponencial. El segundo ejemplo fue utilizado por Dempster et al. (1977) para mostrar el funcionamiento del algoritmo EM. Otro autor que también lo considera con este mismo objetivo es Pawitan (2001, p. 343). El ejemplo en cuestión es el famoso ejemplo genético de Rao (1973, p. 369). Este ejemplo sirve para mostrar también como este algoritmo funciona en problemas que no son de datos incompletos, pero pueden formularse como tales. El tercer ejemplo está basado en una muestra aleatoria de tamaño n de la distribución Gamma con parámetros de forma y escala α y λ respectivamente. Como es conocido, en la estimación máximo verosímil de los parámetros, uno de ellos (α) requiere de la solución de una ecuación no lineal. Todos los cálculos fueron realizados. Todos los cálculos fueron realizados con el paquete Matlab (versión 6.5. Release 13)

2. Ejemplos

Ejemplo 1.

Considérese una muestra aleatoria de tamaño n de una distribución de probabilidad Exponencial con media λ . De dicha muestra solamente se conoce su tamaño y el valor máximo de los valores en la muestra. La función de densidad para este modelo probabilístico está definida por

$$f_X(x; \lambda) = \lambda^{-1} e^{-\lambda^{-1}x}$$

para valores positivos de la variable y y el parámetro. Sea “ y ” el valor máximo de las observaciones x_1, \dots, x_n que conforman la muestra aleatoria de tamaño n de la distribución que se ha adoptado. Es conocido que en el caso de una distribución de probabilidad continua, la ley de distribución del estadístico de orden “máximo de las observaciones”, es caracterizada por la función de densidad

$$f_Y(y; \lambda) = n[F_X(y)]^{n-1}f_X(y; \lambda) \quad , \quad (2.1)$$

donde F_X es la función de distribución de la variable aleatoria Y . De (2.1) se obtiene que la función de verosimilitud está dada por

$$L(\lambda; y) = [F_X(y)]^{n-1}f_X(y; \lambda) \quad (2.2)$$

y de la expresión (2.2) bajo el modelo Exponencial se tiene que la ecuación de verosimilitud a resolver para obtener la estimación máximo verosímil es

$$\frac{\partial}{\partial \lambda} \log L(\lambda; y) = -\frac{(n-1)e^{-\frac{y}{\lambda}}}{\lambda^2(1-e^{-\frac{y}{\lambda}})} - \frac{1}{\lambda} + \frac{y}{\lambda^2} = 0. \quad (2.3)$$

La solución de (2.3) requiere de algún método numérico. En este trabajo se ensaya la aplicación del algoritmo EM por tratarse de un caso típico de datos incompletos y se

compara con la propuesta que se hace en este trabajo, en cuyo caso la función que se evalúa numéricamente y a partir de la cual se obtiene el máximo y el valor del parámetro que lo produce está dada por

$$\log L(\lambda; y) = (n - 1) \log \left(1 - e^{-\frac{y}{\lambda}} \right) - \log \lambda - \frac{y}{\lambda}.$$

Para la comparación se generaron muestras aleatorias de la distribución Exponencial para diferentes valores de n y de λ . En la tabla (2.1) se muestran los resultados obtenidos para algunas de las simulaciones realizadas. En el algoritmo EM como valor inicial para comenzar el proceso iterativo se tomó como valor inicial la tercera parte del máximo de la muestra. En el caso del método que se propone en este trabajo se evaluó el logaritmo de la función de verosimilitud en un conjunto de valores del parámetro, comenzando desde la décima parte del máximo de la muestra hasta el valor de este máximo con incrementos de 0.0001. Empíricamente estos valores demostraron ser adecuados para el estudio que se realizó.

Tabla 2.1

| Tamaño de muestra n | Valor del parámetro λ | λ estimado por método que se propone | λ estimado por algoritmo EM |
|-----------------------|-------------------------------|--|-------------------------------------|
| 5 | 2 | 1.5629 | 1.5629 |
| 20 | 2 | 2.6249 | 2.6249 |
| 5 | 20 | 21.9416 | 21.9415 |
| 20 | 20 | 23.2709 | 23.2709 |

Ejemplo 2

En la situación que se considera ahora se asume un modelo Multinomial con cuatro celdas, del cual se tiene una muestra de tamaño 197 y donde las frecuencias por celdas y las probabilidades de cada una de ellas están dadas en los siguientes vectores $\mathbf{y} = (125, 18, 20, 34)$ y $\boldsymbol{\theta} = \left(\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4} \right)$ respectivamente. En realidad la ecuación de verosimilitud para este caso tiene una solución numérica exacta (0.627) que resulta de la solución de una ecuación de segundo grado. Obviamente esta ecuación también puede resolverse aplicando un método numérico, o el algoritmo EM como ya se mencionó y cuyo resultado es también 0.627. El logaritmo de la función de verosimilitud que se obtiene de este modelo Multinomial está dado por la expresión

$$\log L(\boldsymbol{\theta}; \mathbf{y}) = 125 \log \left(\frac{1}{2} + \frac{\theta}{4} \right) + 38 \log \left(\frac{1-\theta}{4} \right) + 34 \log \left(\frac{\theta}{4} \right). \quad (2.4)$$

Para la evaluación numérica de esta función, para posteriormente determinar su valor máximo y el valor de θ que lo produce, se utilizaron valores comenzando por un valor inicial del parámetro $\theta_0 - 0.10$ y hasta $\theta_0 + 0.10$ con incrementos de 0.001 en una primera etapa. Una vez determinado el valor del parámetro ($\hat{\theta}$) que maximiza a la expresión (2.4), se evalúa ésta nuevamente, pero ahora en un conjunto de valores que comienza en $\hat{\theta} - 0.01$ hasta el valor $\hat{\theta} + 0.01$ con incrementos de 0.0001. El denominado valor inicial del parámetro es una primera estimación a partir de la información de una de las cuatro celdas

y su probabilidad correspondiente. El valor inicial que se utilizó fue el de 0.69, el cual se obtiene de considerar la frecuencia de la celda cuatro y su correspondiente expresión de su probabilidad. Es decir,

$$\theta_0 = \frac{(4 \times 34)}{197} = 0.69$$

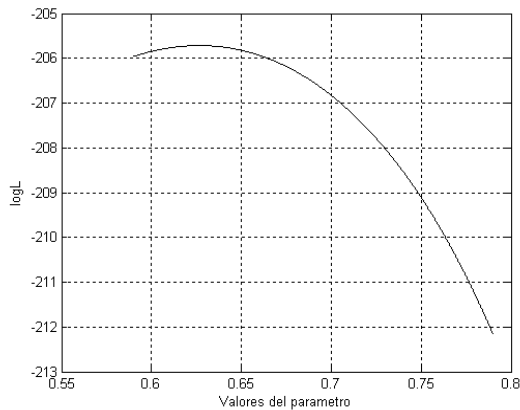
En las gráficas 2.1 y 2.2 se muestran estos dos momentos. El método propuesto que se evalúa arrojó el valor de $\theta = 0.6268$.

Ejemplo 3

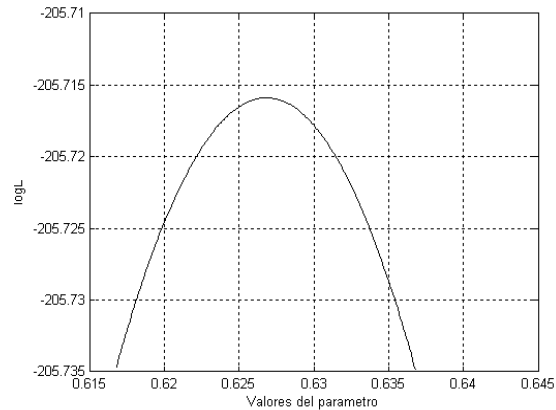
Supóngase una muestra aleatoria de tamaño n de la distribución de probabilidad Gamma con parámetros de forma y escala α y λ respectivamente. Es conocido que la complejidad en la determinación de la estimación máximo verosímil de estos parámetros radica en la expresión de la ecuación producida por

$$\frac{\partial}{\partial \alpha} \log L(\alpha, \lambda; \mathbf{y}) = n \log \left(\frac{\bar{y}}{\alpha} \right) + n \frac{\partial}{\partial \alpha} \log \Gamma(\alpha) - \sum_{i=1}^n \log y_i = 0, \quad (2.5)$$

lo cual requiere de un método numérico en su solución ($\hat{\alpha}$) respecto a α , para después obtener la estimación de λ mediante la expresión $\hat{\lambda} = \frac{\bar{y}}{\hat{\alpha}}$.



Gráfica 2.1



Gráfica 2.2

Para la evaluación del método propuesto se generaron varias muestras aleatorias de diferentes tamaños de la distribución Gamma, encontrándose la estimación máximo verosímil de los parámetros mediante la aplicación del método de Newton para la solución de la ecuación (2.5) y el algoritmo que se propone. Algunos de los resultados obtenidos se muestran en la tabla 2.2. Si se denota la media y la varianza de la distribución Gamma por μ y σ^2 respectivamente, y conocido que $\mu = \alpha\lambda$ y $\sigma^2 = \alpha\lambda^2$, se puede obtener una primera aproximación de la estimación de α y λ , sustituyendo los valores de μ y σ^2 por la media y la varianza muestrales, de tal forma que $\alpha_0 = \frac{\bar{y}^2}{s^2}$ y $\lambda_0 = \frac{\bar{y}}{\alpha_0}$. Para la aplicación del método que se propone se evaluó la función logaritmo de la verosimilitud

$$\log L(\alpha, \lambda; \mathbf{y}) = -n \log \Gamma(\alpha) - n \alpha \log \lambda + (\alpha - 1) \sum_{i=1}^n \log x_i - \lambda^{-1} \sum_{i=1}^n x_i,$$

para valores de α de $\alpha_0 - 0.5$ a $\alpha_0 + 0.5$ y para valores de λ de $\lambda_0 - 0.5$ a $\lambda_0 + 0.5$, con incremento de 0.01 en ambos casos.

Tabla 2.2

| Tamaño de muestra n | Valor del par. α | Valor del par. λ | Estimación Newton α | Estim. Newton λ | Valor de logL | Estim. Método propuesto α | Estim. Método propuesto o λ | Valor de logL |
|------------------------|----------------------------|-----------------------------|-------------------------------|----------------------------|---------------|-------------------------------------|---|---------------|
| 10 | 10 | 5 | 18.8431 | 2.8143 | -39.0379 | 18.8538 | 2.8038 | -39.0297 |
| 20 | 10 | 5 | 10.3591 | 5.8300 | -86.3583 | 10.3588 | 5.8279 | -86.3584 |
| 10 | 20 | 10 | 24.2353 | 7.7751 | -50.4987 | 23.5269 | 7.7514 | -50.5330 |
| 20 | 20 | 10 | 18.2240 | 11.0000 | -104.993 | 18.5088 | 10.8258 | -104.994 |
| 50 | 25 | 30 | 26.4057 | 28.3154 | -319.319 | 26.6633 | 28.0283 | -319.247 |

3. Conclusiones

Por los resultados obtenidos en las tres diferentes situaciones donde se ha evaluado el desempeño del método que se ha propuesto, hay evidencia de que el mismo se desempeña bien, en particular en los dos primeros ejemplos, donde el objeto de estimación era un solo parámetro. En el tercer ejemplo, donde el objeto de estimación es de dos parámetros, no hay tanta coincidencia como en el caso de los dos ejemplos anteriores, aunque los resultados son aceptables. Esto puede ser motivado por el incremento de los valores en el rango que se considera, ya que de haberse fijado el incremento en 0.001, el algoritmo que se utiliza se vuelve ineficiente por el tiempo que emplea en evaluar la función que se desea maximizar.

REFERENCIAS

- Dempster, A., Laird, N.M., and Rubin, D. B. (1977). Maximun likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, 39, 1-38.
- MATLAB Versión 6.5 (2002). *The MathWorks, Inc.* www.mathworks.com.
- McLachlan, G. J., and Krishnan, T. (1997). *The EM algorithm and extensions*. New York. Wiley
- Pawitan, Y. (2001). *In all likelihood*. New York: Oxford University Press.
- Quintana, P., Villalobos, E. y Cornejo, M. C. (2005). *Métodos numéricos con aplicaciones en Excel*. Editorial Reverté. México.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. 2nd. edn. New York. Wiley.

Sobre la Potencia Prospectiva: un Programa para su Cálculo

Luis Omar Jamed¹, Ernesto Menéndez Acuña²

¹Facultad de Ingeniería Química. Universidad Veracruzana. lojb33@gmail.com

²Facultad de Matemáticas. Universidad Veracruzana. emenendez@gmail.com

1. Introducción

En este trabajo se presenta un programa que permite evaluar los valores de los elementos que conforman el estudio prospectivo de la potencia de una dócima o prueba de hipótesis en el marco del análisis de varianza, desde una perspectiva de su utilidad en la fase de la planeación del experimento a realizarse. En este contexto se acostumbra denominar a la potencia con el adjetivo de prospectiva; entendiéndose el término de potencia prospectiva como el cálculo de la probabilidad previa a la experimentación para detectar diferencias significativas entre los tratamientos que se someterán a prueba, cuando estas diferencias existen.

La teoría básica del cálculo prospectivo de la potencia fue desarrollada en la década de los años 50's y fundamentada por Scheffé (1959). Actualmente, a pesar del desarrollo de los medios computacionales y que el concepto de potencia ha sido tratado en detalle por varios autores, tales como: Kraemer and Thiemann (1987), Cohen (1988), Lipsey (1990) y Muller and Benignus (1992), por solo citar algunos, el desconocimiento de tal concepto y su utilidad desde la perspectiva prospectiva es bastante generalizado y en el mejor de los casos, cuando el investigador conoce la importancia y utilidad de la misma, puede no sentirse motivado a usarla por las complicaciones de cálculo que pudieran derivarse.

Existen muchos programas estadísticos, tanto comerciales como libres. Algunos de ellos incluyen en el módulo de diseño de experimentos, el cálculo de la potencia. No obstante lo anterior, lo hacen desde la perspectiva de la potencia retrospectiva. Entendiéndose por "retrospectiva" como el cálculo de la potencia una vez concluido el análisis de los datos. En consecuencia, los resultados de estos programas en la etapa de planificación son bastante limitados. Esta insuficiencia no es difícil de entender a partir de que dichos programas no están pensados en función del uso de la potencia en la etapa de la planificación del diseño previa a la experimentación.

Los potentes programas comerciales de aplicación al diseño de experimento, tales como: el JMP(V5©2002, SAS Institute Inc.) y el Design-Expert(V7.1.1©2007, Stat-Ease Inc.) no brindan la posibilidad del cálculo prospectivo de la potencia de una manera amigable. El JMP por ejemplo, si se deseara hacer análisis de la potencia previa a la experimentación, requiere programación en su lenguaje y conocimiento de la teoría de análisis prospectivo de la potencia. El Design-Expert(V7.1.1©2007, Stat-Ease Inc.), lo hace de una manera tal, que no permite la consideración de otros valores de diferencias significativas, sino aquellas preestablecidas en el programa, las cuales son a lo más tres. Como consecuencia, no permiten un análisis exploratorio exhaustivo de la potencia prospectiva, por lo cual el

investigador no puede valorar diferentes alternativas y con ello tomar la decisión más conveniente, en función de sus objetivos, expectativas y recursos.

En este trabajo, se ofrece un programa el cual permite el cálculo y análisis de la potencia prospectiva bajo un amplio escenario de condiciones, de manera fácil y amigable, incluso con salidas gráficas. Si bien este programa está dirigido al caso de los diseños experimentales factoriales, la flexibilidad que se ha logrado en su construcción, permite el estudio prospectivo de la potencia en otros tipos de diseños. Esta flexibilidad del programa es posible por su estructura en módulos. El programa se elaboró mediante la codificación en Matlab V7.2.0.232 (MathWorks Inc., 2006). Con el uso de este programa, el investigador está capacitado para seleccionar, en la fase de planificación del experimento, los valores de compromiso en los elementos que conforman el estudio prospectivo de la potencia.

La potencia prospectiva contiene cinco elementos que intervienen en su cálculo (Menéndez y Jamed, 2006), estos elementos son: (a) el tamaño de efecto Δ a ser detectado, (b) la variación inherente σ^2 de la variable que se observa, (c) el número n de réplicas en cada combinación de niveles de los factores, (d) la significación α de la prueba estadística y (e) la magnitud $(1 - \beta)$ mínima deseable de la potencia.

Para ejemplificar el procedimiento y el uso del programa se presentan los siguientes diseños.

1. Diseño de clasificación simple.
2. Diseño factorial completo.
3. Diseño “Box–Wilson” de composición central.

2. Procedimiento general de utilización de la potencia en la planificación del diseño de experimento.

- a. Seleccionar el diseño experimental de interés.
- b. Obtener los valores de los elementos que se encuentran asociados en la potencia prospectiva (Δ , σ o $(k = \Delta/\sigma)$, α , n y $(1 - \beta)$). Ya que en la planificación habitualmente se desconoce σ , se recomienda la utilización de k . Siendo fijo el nivel de significación α , los elementos asociados son: k , n y $(1 - \beta)$. En este punto, es importante el cálculo de los valores asociados k , n y $(1 - \beta)$ que brinda el programa de cómputo.
- c. Tomar la decisión de la terna correspondiente k , n y $(1 - \beta)$. En este punto, el investigador evalúa la terna más adecuada conforme a su disponibilidad de recursos y alcance buscado en la investigación.
- d. Proceder a la etapa de experimentación.

3. Ejemplos

- 1.1 Utilización de la potencia prospectiva en el diseño de clasificación simple.

El diseño contiene un factor de cuatro niveles o tratamientos con significación de $\alpha = 0.05$. Por consideraciones prácticas, sólo es posible considerar la posibilidad de 3 o 4 réplicas en cada nivel, determinar el tamaño de efecto estandarizado k , tal que cumpla con una potencia prospectiva de 0.80. Para $n = 3$ (no. de réplicas en cada nivel), los grados de libertad son: $\nu_{total} = 11$, $\nu_{factor} = 3$, $\nu_{res} = 8$. Para $n = 4$, los grados de libertad son: $\nu_{total} = 15$, $\nu_{factor} = 3$, $\nu_{res} = 12$. En la tabla 1.1 se presentan los valores de la potencia prospectiva y en las figuras 1.1 y 1.2 se presentan sus gráficas.

Tabla 1.1. Potencia prospectiva a $n = 3, 4$

| k | n = 3 | n = 4 |
|------|----------|----------|
| 0.00 | 0.050000 | 0.050000 |
| 0.50 | 0.064635 | 0.072664 |
| 1.00 | 0.112979 | 0.150728 |
| 1.50 | 0.203798 | 0.298955 |
| 2.00 | 0.339058 | 0.503705 |
| 2.50 | 0.504399 | 0.712094 |
| 2.80 | 0.606276 | 0.814069 |
| 3.00 | 0.671001 | 0.867819 |
| 3.50 | 0.809582 | 0.953100 |
| 4.00 | 0.904934 | 0.987333 |
| 4.50 | 0.959345 | 0.997419 |
| 5.00 | 0.985172 | 0.999605 |

Figura 1. Potencia prospectiva a $n = 3$

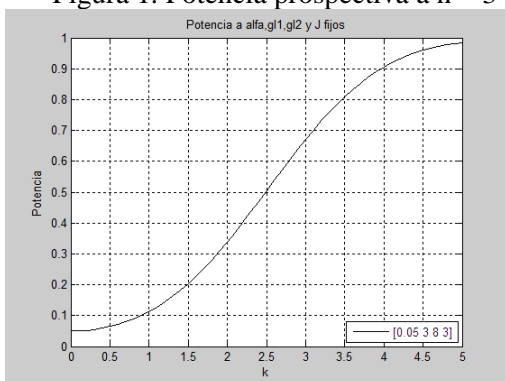
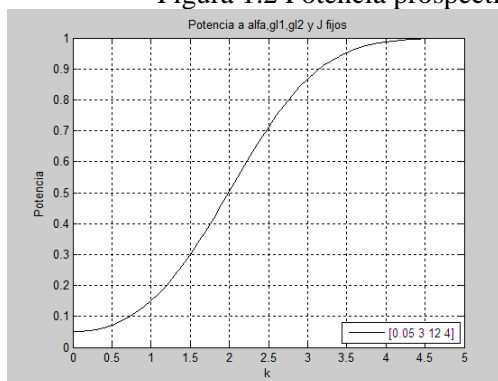


Figura 1.2 Potencia prospectiva a $n = 4$



En la tabla 1.1 se observa para $n = 3$, una potencia de 0.809582 con un efecto estandarizado $k = 3.50$ y para $n = 4$, una potencia de 0.814069 con un efecto estandarizado $k = 2.80$. Lo anterior se interpreta de la siguiente manera. Se espera en el 80% de los casos que la prueba de hipótesis sea significativa, detectar un tamaño de efecto $\Delta = 3.5\sigma$ cuando se utilicen 3 réplicas. Se espera en el 81% de los casos que la prueba de hipótesis sea significativa, detectar un tamaño de efecto $\Delta = 2.8\sigma$ cuando se utilicen 4 réplicas.

1.2 Utilización de la potencia prospectiva en el diseño factorial completo.

Se desea obtener los valores de la potencia del tratamiento interacción triple ($m = 3$) para un valor fijo del tamaño de efecto $\Delta = 2\sigma$ ($k = 2$). El diseño factorial es un 3^3 , considerando número de réplicas desde 2 hasta 5 en cada condición, con significación del

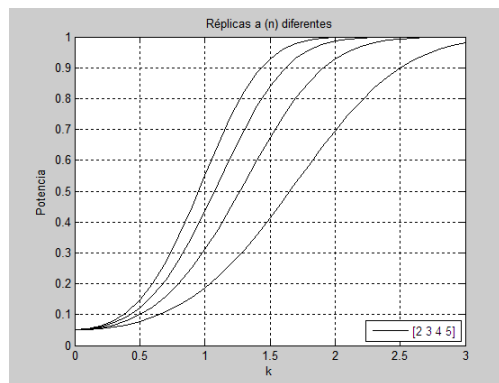
5%. En la siguiente tabla 2.1, se presentan los valores de la potencia prospectiva y en la figura 2.1 se presenta su gráfica.

En la tabla 2.1 se observa en un tamaño de efecto estandarizado $k = 2$ y siendo significativo el efecto de interacción triple, lo siguiente. Con $n = 2$, se esperan detectar el 69.53% de los casos. Con $n = 3$, se esperan detectar el 92.81% de los casos. Con $n = 4$, se esperan detectar el 98.68% de los casos. Y por último, con $n = 5$, se esperan detectar el 99.80% de los casos.

Tabla 2.1. Potencia del efecto interacción triple, diseño 3^3 , con diferentes réplicas $n = 2, 3, 4, 5$

| k | n = 2 | n = 3 | n = 4 | n = 5 |
|----------|--------------|--------------|--------------|--------------|
| 0.00 | 0.050000 | 0.050000 | 0.050000 | 0.050000 |
| 0.50 | 0.077646 | 0.099630 | 0.122736 | 0.147289 |
| 1.00 | 0.186668 | 0.310491 | 0.434725 | 0.551415 |
| 1.50 | 0.412364 | 0.673801 | 0.840093 | 0.929404 |
| 2.00 | 0.695332 | 0.928115 | 0.986806 | 0.998006 |
| 2.50 | 0.899160 | 0.993925 | 0.999749 | 0.999992 |
| 3.00 | 0.980372 | 0.999824 | 0.999999 | 1.000000 |

Figura 2.1 Potencia del efecto interacción triple, diseño 3^3 , con diferentes réplicas $n = 2, 3, 4, 5$



1.3 La potencia prospectiva en el diseño “Box–Wilson” de composición central.

Se desea calcular la potencia prospectiva de un diseño de composición central de segundo orden, el cual contiene 3 factores, 2 réplicas en cada punto factorial, 8 réplicas en el punto central y 2 réplicas en cada punto axial. Se desea también asegurar que los efectos cuadráticos contenidos en el modelo no se encuentren correlacionados. Conforme a lo anterior, el diseño seleccionado debe ser de tipo de distancia axial ortogonal.

Identificar los tamaños de efecto estandarizados que cumplen obtener al menos una potencia de 0.80 en cada efecto en estudio (modelo, interacción, lineal y cuadrático).

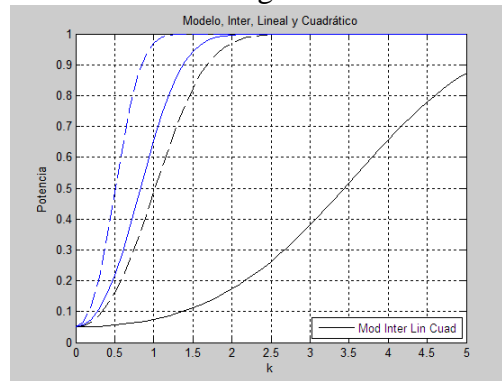
La tabla 3.1 muestra los valores más importantes de la potencia prospectiva, la figura 3.1 representan sus gráficas.

En la tabla 3.1, se observan los siguientes valores del efecto estandarizado k. Para el efecto conjunto del modelo, $k \geq 4.60$. Para el efecto interacción, $k \geq 1.50$. Para el efecto lineal, $k \geq 1.20$. Para el efecto cuadrático, $k \geq 0.75$. Los valores anteriores de k corresponden a mantener al menos una potencia de 0.80.

Tabla 3.1

| k | modelo | inter | lineal | cuad |
|------|----------|----------|----------|----------|
| 0.00 | 0.050000 | 0.050000 | 0.050000 | 0.050000 |
| 0.50 | 0.055830 | 0.161219 | 0.218494 | 0.486480 |
| 0.75 | 0.063509 | 0.303545 | 0.424460 | 0.823196 |
| 1.00 | 0.074972 | 0.486480 | 0.654815 | 0.970511 |
| 1.20 | 0.087309 | 0.637117 | 0.807673 | 0.996050 |
| 1.50 | 0.112018 | 0.823196 | 0.942427 | 0.999930 |
| 2.00 | 0.173087 | 0.970511 | 0.997032 | 1.000000 |
| 2.50 | 0.262861 | 0.997802 | 0.999957 | 1.000000 |
| 3.00 | 0.380562 | 0.999930 | 1.000000 | 1.000000 |
| 3.50 | 0.517060 | 0.999999 | 1.000000 | 1.000000 |
| 4.00 | 0.655902 | 1.000000 | 1.000000 | 1.000000 |
| 4.50 | 0.778915 | 1.000000 | 1.000000 | 1.000000 |
| 4.60 | 0.800336 | 1.000000 | 1.000000 | 1.000000 |
| 5.00 | 0.873340 | 1.000000 | 1.000000 | 1.000000 |

Figura 3.1



Identificación, trazos de efectos
 — Modelo - - - - Interacción
 — Lineal - - - - Cuadrático

Con toda seguridad con los resultados de este trabajo se contribuye a fomentar la cultura estadística alrededor de la temática de la potencia prospectiva, favoreciendo tanto su aplicación como su enseñanza.

REFERENCIAS

- Cohen, J.(1988). “Statistical Power Analysis for the Behavioral Sciences”, 2nd Edition. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.
- Design-Expert(V7.1.1©2007, Stat-Ease Inc.), (Feb 12, 2007). Stat-Ease, Inc. 2021 East Hennepin Ave., Suite 480. Minneapolis, Mn 55413. www.statease.com.
- JMP(V5©2002, SAS Institute Inc.), SAS Institute Inc. www.jmpdiscovery.com.
- Kraemer H.C. and Thieman S.(1987). “How Many Subjects?”, Sage Publications, London, UK.
- Lipsey, M.W.(1990). “Design Sensitivity: Statistical Power for Experimental Research”. Sage, Newbury Park, California.
- MATLAB Version 7.2.0.232(2006a): The MathWorks, Inc. www.mathworks.com.
- Menéndez E., Jamed L.O.(2006). “Una propuesta para el cálculo de la potencia en el ANOVA”. Revista Investigación Operacional Vol. 27 No. 2. pp. 194-205. Ernesto Menéndez Acuña (Cuba) email: ema@matcom.uh.cu, Luis O. Jamed Boza (México) email: lojb@prodigy.net.mx.
- Muller, K.E., and Benignus, V.A.(1992). “Increasing scientific power with statistical power”. Neurotoxicology and Teratology **14**:211-219.
- Scheffé, H. (1959). “The Analysis of Variance”. John Wiley & Sons. New York.

Contribución de la Psicología en el Desarrollo de la Estadística

Mtro. José García Franco

Facultad de Estudios Superiores Zaragoza, UNAM. jgfranco56@yahoo.com

La enseñanza de la estadística en las escuelas de psicología en el Distrito Federal, desde hace 36 años, ha contenido una serie de errores, dentro de los cuales cabe destacar, la enseñanza de la estadística como una materia sin relación, no sólo con la psicología, sino además con la metodología de investigación. Situación que se deja a la inteligencia y al insight del alumno, el cual debe integrarla con la psicología y la investigación, sin la ayuda de la escuela. Esto significa que no existe un plan curricular que permita esta fusión, desde la aparición de la Facultad de Psicología de la UNAM en 1973 a la fecha. Más aun si se considera que el modelo curricular de ésta, es ejemplo conductor para las demás escuelas de psicología del Distrito Federal en particular y del país en general. Otro error ha consistido en la impartición de la materia por matemáticos o ingenieros, lo cual genera una enseñanza con ejemplos propios de esas disciplinas y por lo tanto ajenos a la psicología. Uno más de los errores ha sido considerar a la estadística totalmente ajena y sin alguna relación con la psicología. Todo esto ha generado un miedo considerable en los alumnos, además de un alto índice de reprobación. Por esta razón y por experiencia propia, se pretende en este trabajo contribuir al conocimiento de la historia de la estadística y considerar la participación de eminentes psicólogos en el desarrollo de ésta, para así generar en los alumnos una conciencia de que la estadística no es una materia ajena o desarrollada sólo por matemáticos, que tiene que ser cursada como un requisito sin conexión con su vida profesional, sino como una disciplina propia y necesaria para su desarrollo profesional.



En estos tiempos además, el contar con internet permite a académicos y alumnos conocer valiosa información, por ejemplo acceder a biografías de hombres que han dedicado su vida al desarrollo del conocimiento, en este caso el de la estadística, así que vaya este humilde reconocimiento a esos hombres y sirva de modesto ejemplo de la información que navega en la web.

La contribución de la psicología en el desarrollo de la estadística inicia en Alemania con Gustav Fechner, que contaba con estudios incompletos en medicina, incursiona en las técnicas de la experimentación, para describir la relación entre estímulos y sensación y deriva la Estadística hacia el campo de la psicología experimental. Fechner introdujo la medición en la psicología. Aparentemente creó el término psicofísica para describir la psicología experimental, que practicó como disciplina formal, a mediados del siglo XIX.

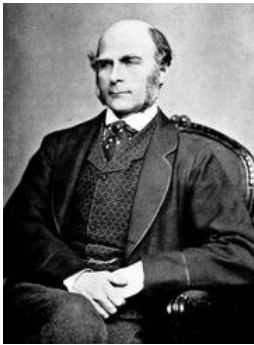


http://en.wikipedia.org/wiki/Gustav_Fechner

Posteriormente aparece en escena otro psicólogo, Hermann Ebbinghaus, que siguió la línea de Fechner y compartió dos ideas

cruciales con él: Que el estudio cuantitativo era la única manera de expresar con precisión las vagas nociones que la psicología manejaba antes, y el convencimiento que el apoyo de un cuidadoso diseño experimental es fundamental para la experimentación. Ebbinghaus aplicó estas ideas al estudio de la memoria. Vale la pena mencionar que al no contar con sujetos de experimentación, él mismo fue su sujeto. A él se debe la elaboración de la curva del olvido.

A partir de 1880, dos hombres, Francis Galton y Karl Pearson, crean una revolución en la Estadística, al proporcionar una metodología empírica que sustituye a la experimentación controlada, en disciplinas donde la experimentación no es posible de aplicar. Esta metodología empírica ya había sido utilizada en la psicología.



Francis Galton, primo de Darwin, nació en 1822 y a pesar de no ser psicólogo, investigó el carácter hereditario de la genialidad o inteligencia, utilizando curvas normales inversas, que llamó "ojivas", término que tomó de la arquitectura, y que aún se utiliza. Fue pionero en el tema de la regresión lineal simple, o reversión, como él la llamó, técnica para obtener una expresión que relaciona en forma lineal, dos características. También se preocupó de la estimación de los componentes de la varianza, o partes de la variabilidad de un fenómeno observado, atribuibles a causas identificables. El concepto estadístico más conocido de él, es el de la correlación, aún cuando su interpretación es limitada a un coeficiente que mide el grado de asociación entre el comportamiento de dos variables.

http://www.jorgegalbiati.cl/ejercicios_4/HistoriaEstadistica.pdf

La pregunta más importante que generó fue ¿Se podrán medir las capacidades mentales? Con ello se convirtió en uno de los padres de la Psicometría, conjuntamente con Alfred Binet en Francia y Catell en Norteamérica, este último alumno de doctorado de Wilhelm Wundt, Padre de la Psicología experimental (Morales, 1976). Sin embargo quizá la mejor contribución de Sir Francis Galton fue la de ser el protector de un hombre genial: Karl Pearson.

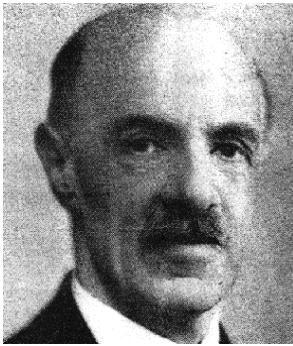


Karl Pearson aunque no fue psicólogo siempre desarrolló una gran fascinación por todas las ciencias en especial las ciencias sociales. Él mostró interés en distribuciones probabilísticas asimétricas, en contraposición con las distribuciones normales, simétricas. Llegó de esta manera a introducir una familia de distribuciones probabilísticas, hoy conocida como Gama, que ya había sido descubierta independientemente por el estadounidense Erastus De Forest.

Pearson mostró interés en los más diversos temas, además de la estadística, como la filosofía, la religión, la historia, entre otros. Su "Gramática de las Ciencias", de 1892, ilustra su convicción de que la estadística analítica yace en los fundamentos de todo el conocimiento.

Pearson, en su trabajo, dio más importancia a la cuantificación de la correlación entre dos variables, en la forma de un coeficiente, que la que le había dado Galton. Él y otros investigadores desarrollaron varios coeficientes de correlación, para el estudio de diferentes problemas en genética, biología, y otras disciplinas. El más común y conocido de ellos, hoy en día, lleva su nombre. A Karl Pearson se debe, también, la prueba estadística ji cuadrada, introducida en 1900. Este estadístico, es utilizado como medida de asociación entre dos tablas de frecuencia y una de sus aplicaciones es el probar el ajuste de una ley probabilística a un conjunto de datos empíricos.

Con posterioridad al desarrollo del concepto de correlación, de Galton y de Pearson, problemas de clasificación en antropología y botánica dieron origen a coeficientes de similitud y a las funciones discriminantes del análisis multivariante. Dentro de esta misma rama de la estadística, pero en otra dirección, el estudio de las respuestas de los tests mentales dio origen a técnicas de reducción de dimensionalidad, es decir, el reemplazo de un gran número de variables correlacionadas, por un pequeño grupo de variables construidas a partir de las primeras, que contienen aproximadamente la misma cantidad de información. Entre ellas se encuentra el análisis factorial, que permite encontrar y cuantificar factores que influyen sobre las respuestas observadas. Las raíces del análisis factorial se encuentran en la psicología.



Charles Spearman condiscípulo de Catell y también alumno de doctorado de Wundt, nació en 1863 en Inglaterra. Spearman sirvió en el ejército inglés, participando en la guerra de los Boers. Luego se retiró, para dedicarse a estudiar psicología, llegando a desempeñarse, ya tarde en su vida, como profesor de psicología en el University College, de Londres. Es conocido por sus contribuciones al análisis factorial, que se mencionó como una de las técnicas de la rama de la Estadística denominada análisis multivariante. También es conocido por sus investigaciones sobre la inteligencia. Estos intereses lo obligaron a estudiar estadística,

llevándolo a desarrollar un coeficiente de correlación basado en rangos, que hoy lleva su nombre. El trabajo de Spearman ha sido desarrollado con posterioridad, desembocando en el análisis de varianza multivariante.

http://www.jorgegalbiati.cl/ejercicios_4/HistoriaEstadistica.pdf

Al igual que Spearman, otros psicólogos han entregado importantes contribuciones a la Estadística, lo que es indicación de su gran utilidad en el estudio de los fenómenos



psíquicos. Entre ellos se cuentan L.L. Thurstone y R. Likert, quienes se preocuparon de problemas de escalamiento, que consiste en transformar una característica medida en una escala conceptual, a una escala numérica. El escalamiento está fuertemente ligado al diseño y análisis de encuestas y tests.

Louis Thurstone fue un psicólogo norteamericano que comenzó trabajando en ingeniería electrónica antes de iniciar sus estudios de Psicología. Estudió en las universidades de Cornell y Chicago. Después de enseñar psicología en el Instituto Carnegie de Tecnología

(1915-23), fue psicólogo del Instituto Gubernamental de Investigaciones (1923-24) y profesor en la Universidad de Chicago (1924-52), donde en 1938 ocupó la cátedra Charles F. Grey. Inventó una escala psicofísica para la medida de las actitudes por el método de intervalos de aparentemente iguales e introdujo nuevas técnicas en el campo del análisis de factores. Dio un gran impulso a la Psicometría. Se opuso al factor G de Spearman y postuló sus siete aptitudes mentales primarias: comprensión verbal, comprensión espacial, razonamiento abstracto, capacidad numérica, fluidez verbal, capacidad perceptiva y memoria. Escribió: Un estudio Factorial de la Percepción en (1944) y Análisis de Múltiples Factores en 1947.

<http://www.biografiasyvidas.com/biografia/t/thurstone.htm>



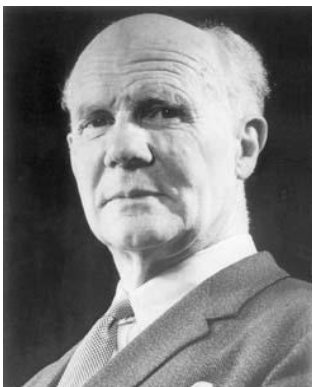
Rensis Likert (1903-1981) Psicólogo norteamericano se graduó en la Universidad de Michigan en economía y sociología en 1922. Recibió su doctorado en Psicología en la universidad de Columbia en 1932. Fue el director fundador del Instituto de Investigación Social en 1949.

Trabajó en un buen número de agencias gubernamentales en el negocio de obtención de información valiosa y pasó gran parte de su vida perfeccionando y desarrollando un método confiable para obtener datos de los entrevistados. Este trabajo se publicó en 1932 con el título de, “Una técnica para medir actitudes”. Sus ideas han sido una gran influencia en el siglo veinte y seguramente también

en el veintiuno. Trabajó en empresas y desarrolló su teoría organizacional, la cual fue bien acogida en Japón.

<http://www.answers.com/topic/rensis-likert>

Egon Pearson, hijo de Karl Pearson y Neyman, presentaron en 1936 una teoría sobre la forma de probar hipótesis estadísticas, con base en datos. Esta presentación promovió mucho interés, estimuló una considerable cantidad de investigación, y muchos de los resultados hasta hoy aún se usan. Ellos resolvieron dificultades fundamentales para la comprensión de las pruebas de hipótesis, introduciendo las nociones de hipótesis alterna, y los dos tipos de error, el de rechazar una hipótesis que es verdadera, y el de no rechazar una hipótesis que es falsa.



Egon Pearson, aunque tampoco fue psicólogo, sino astrónomo, fue también como su padre un gran estadista y lo sucedió como profesor de estadística en la University College London y como editor de la revista Biométrica. Él fue presidente de la [Royal Statistical Society](#) de 1955 a 1956.

<http://www.statemaster.com/encyclopedia/Egon-Pearson>

Al conocer a estos grandes hombres que han construido la actual fisonomía de la Psicología y su enorme contribución en la Estadística y por supuesto a la conformación de la investigación, tal como hoy se trabaja, no puede uno sustraerse a la plena conciencia de que la estadística no debe resultar ajena, sino por el contrario, es como llegar a un buen refugio, es como llegar a casa. Por esta razón dentro de la enseñanza de la estadística es necesario partir no de números y fórmulas, sino de la construcción del conocimiento que da el sentido de pertenencia y el amor por la ciencia, la comprobación y la verdad.

Bibliografía

Galbiati, J. *Desarrollo histórico de la estadística*.
http://www.jorgegalbiati.cl/ejercicios_4/HistoriaEstadistica.pdf [Consulta: martes, 16 de Junio de 2009]

Morales, M.L. (1976) *Psicometría Aplicada*. México: Ed. Trillas.

<http://www.biografiasyvidas.com/biografia/t/thurstone.htm> [Consulta: miércoles, 17 de Junio de 2009]

<http://www.answers.com/topic/rensis-likert> [Consulta: jueves, 18 de Junio de 2009]

<http://www.statemaster.com/encyclopedia/Egon-Pearson> [Consulta: jueves, 18 de Junio de 2009]

http://en.wikipedia.org/wiki/Gustav_Fechner_ [Consulta: miércoles, 17 de Junio de 2009]

Bibliografía Recomendada

Owen D.B. (1976). *On the history of statistics and probability*. Ed. Marcel Decker.

Stiegler S. (1986) *The history of Statistics*. Ed. Harvad University Press.

Actitud hacia la Estadística en Estudiantes de Estadística e Informática

José Luis Colorado Hernández¹, Sergio Francisco Juárez Cerrillo²

¹Instituto de Psicología y Educación, U. V. icolorado@uv.mx

²Fac. de Estadística e Informática, U. V. sejuarez@uv.mx

Resumen

Se presentan los resultados de la aplicación de un instrumento para evaluar la “*Actitud hacia la Estadística*” en una muestra de estudiantes de las licenciaturas en Estadística e Informática, de la Universidad Veracruzana. El Objetivo del estudio fue caracterizar y comparar la *Actitud hacia la Estadística*, de estudiantes de las licenciaturas de Estadística e Informática. El instrumento utilizado es una Escala tipo Lickert, con 25 ítems, distribuidos en tres dimensiones, *Afectivo*, *Cognitivo* y *Comportamental*. La muestra fue de tipo casual, obtenida en el inmueble donde asisten a clases regularmente los estudiantes de los programas de licenciatura considerados, en un mismo día normal de clases. Los resultados muestran diferencias por grupos de las licenciaturas, en los indicadores considerados en el instrumento utilizado.

Introducción

La problemática del proceso enseñanza de la Estadística es un tópico amplio y sumamente importante dada la importancia de la formación en esta área, “(...) *es frecuente observar la preocupación de alumnos y profesores por el rendimiento inadecuado y por el rechazo a la asignatura de Matemáticas, dentro de la que se incluye a la Estadística.*” (Bazán y Aparicio, 2006 pág. 1). Y esto se debe a la gran importancia que tienen los conocimientos, habilidades y competencias relacionadas con lo cuantitativo en la formación de los estudiantes de los diferentes niveles educativos. “*Tanto la matemática como la Estadística, en el mundo de hoy, pueden considerarse como una herramienta que puede ser utilizada en la búsqueda de resultados y soluciones, y como un estilo de pensamiento que guía la actividad en sus diferentes formas.*” (Bazán y Aparicio, 2006 pág. 1).

Actitudes

En relación a la importancia de las actitudes, Oskamp, y Schultz, (2004), refiere algunas consideraciones al respecto, y Haddock & Maio, (2004) refieren su importancia en el estudio de diferentes necesidades psicológicas, las cuales pueden estar relacionadas con información afectiva, cognitiva, y conductual; además de que permiten predecir el comportamiento de las personas y lo relacionado con el cambio de las actitudes. “(Haddock & Maio, 2004, p. 1).

Genéricamente se mencionan componentes muy semejantes en la conceptualización de las actitudes, “(...) *consideramos que las actitudes siempre se expresan positiva o negativamente, que pueden representar sentimientos vinculados externamente con la asignatura (profesor, actividad, libro, etcétera) y que no se restringen a la afectividad, sino*

que involucran pensamientos, evaluaciones, valoraciones y disposiciones a la acción que forman parte de otros componentes de la personalidad, tal y como veremos después.” (Bazán y Aparicio 2006, pág. 5).

Entre las muchas definiciones de actitudes mencionamos dos que parecen considerar aspectos relevantes en la mayoría de los estudios consultados.

“(…) La actitud constituiría una cierta predisposición o no del sujeto para con algo; una aceptación o no; una cierta manifestación positiva o negativa; una aproximación o alejamiento; un sentimiento favorable o desfavorable. (Bazán y Aparicio 2006, pág. 6).

“(…) una actitud es una predisposición aprendida para responder de manera consistente, favorable o desfavorable, hacia un objeto y sus símbolos. Una actitud tiene dirección: positiva o negativa; intensidad: alta o baja; está conformada por varios elementos tales como: cogniciones o creencias, sentimientos o afectos asociados a evaluaciones, tendencias de comportamiento y se forma principalmente, mediante las experiencias e inferencias o generalizaciones y con base en principios de aprendizaje.” (Ursini, Sánchez y Orendain, 2004, pág. 61).

El estudio de las actitudes hacia la Estadística.

El estudio de las actitudes hacia la Estadística ha sido un tópico de interés en la comunidad académica, considerado de gran relevancia en la medida en que puede ser uno de los principales factores relacionados con la problemática en el proceso de enseñanza aprendizaje de la materia.

“(…) Student attitudes about statistics are important because they may influence the learning process. In particular, student attitudes and beliefs about statistics can affect the extent to which students will develop useful statistical thinking skills, whether they will apply what they have learned outside of the classroom, and whether or not students will choose to enroll in further statistics courses (Gal, Ginsburg, & Schau, 1997). Negative student attitudes toward statistics may create a major obstacle for effective learning (Cashin & Elmore, 1997; Fullerton & Umphrey, 2001; Schutz, Drogosz, White, & Distefano, 1998; Waters, Matelli, Zakrajsek, & Popovich, 1988).” (Mills, J. D. (2004), pag. 351).

La definición del constructo, como es muy frecuente en la investigación educativa, se ha desarrollado en la medida en que se han realizado estudios sobre la temática. Al respecto existen importantes contribuciones en el contexto de la consideración de los tres componentes básicos de las actitudes según refieren los teóricos del asunto. Al respecto Estrada y Batanero y Fortuny, (2004), presentan una interesante propuesta en el contexto de la formación docente.

Después de considerar los anteriores aspectos consideramos que la conceptualización de Ursini y col. (2004), resulta aceptable y considera en su mayoría los diferentes componentes mencionados por los autores anteriormente mencionados, la cual puede ser orientada hacia el objeto “Estadística” en la percepción del alumno.

Así que podemos considerar que la actitud hacia la Estadística muy bien puede ser adaptada en la propuesta de concepto general de actitudes de Ursini y cols. como sigue: *“La actitud hacia la Estadística puede ser considerada como la predisposición aprendida para responder de manera consistente, favorable o desfavorable, hacia la Estadística y sus símbolos, con dirección positiva o negativa, e intensidad alta o baja; la cual está conformada por cogniciones o creencias, sentimientos o afectos asociados a evaluaciones, tendencias de comportamiento y se forma en el sujeto, principalmente mediante las experiencias e inferencias o generalizaciones con base en principios de aprendizaje.”*

En este punto debemos aclarar que nuestro trabajo sobre las actitudes hacia la Estadística se encuentra en fase inicial y para este reporte usaremos los conceptos e instrumento de Estrada, batanero y Fortuny, (2004), el cual se describe más adelante, dejando para el proceso subsiguiente, de esta línea de trabajo, la posible operacionalización y diseño de un instrumento para evaluar la actitud hacia las matemáticas con la definición basada en la conceptualización de Ursini y cols.

Metodología

(a) Objetivos:

Caracterizar y comparar las actitudes hacia la Estadística, con base en el instrumento propuesto por Estrada y cols. (2004), en una muestra de estudiantes de los programas de licenciatura en Estadística y en Informática de la Universidad Veracruzana.

(b) Variables:

Variable dependiente: Actitud hacia la Estadística en sus componentes: Afectivo, Cognitivo y Comportamental.

Variables independientes: Programa de Licenciatura.

(c) Muestra:

Se utilizó una muestra casual de 78 estudiantes de los programas de licenciatura en Estadística e Informática de la Universidad Veracruzana, quienes asistieron a clases durante el periodo semestral Febrero – Julio del 2008.

(d) Instrumento

La escala que utilizamos en este estudio fue diseñada por Estrada y cols. (2004), tomando en cuenta los tres los tres componentes de interés que tradicionalmente se consideran en las escalas de actitudes. Los cuales en la conceptualización de los autores antes mencionados se constituyen en:“(…) Componente cognitivo: Se refiere a las expresiones de pensamiento, concepciones y creencias, acerca del objeto actitudinal, en este caso, la estadística. Incluye desde los procesos perceptivos simples, hasta los cognitivos más complejos. Componente afectivo o emocional: Está constituido por expresiones de sentimiento hacia el objeto de referencia. Recogería todas aquellas emociones y sentimientos que despierta la estadística, y por ello son reacciones subjetivas positivas/negativas, acercamiento/huida, placer/dolor. Componente conductual o tendencial: Aparece vinculado a las actuaciones en relación con el objeto de las actitudes. Son expresiones de acción o intención conductista o conductual y representan la tendencia a resolverse en la acción de una manera determinada.” (Estrada y cols. 2004, pág. 264).

El instrumento construido por Estrada y cols. (2004); se obtuvo combinando las tres escalas siguientes: *escala SAS* (Roberts y Bilderback, 1980); *escala ATS* (Wise, 1985) y la de Auzmendi (1992), y siguiendo las recomendaciones de Osterlind (1989) y Thorndike (1989). Para su elaboración los autores delimitaron el contenido a evaluar y especificaron el formato de los ítems lo cuales constan de un enunciado y una escala de 5 puntos, que valoran las respuestas desde «muy en desacuerdo» (1 punto) hasta «muy de acuerdo» (5 puntos). Refieren los autores la creación de una primera lista de ítems; seguido de una selección contemplando los diferentes componentes, dando un peso equivalente a cada uno. De esta manera, refieren, generaron una lista de 36 enunciados que sometieron a un grupo de expertos para evaluar la adecuación y univocidad de las proposiciones, quedando después de la valoración la escala definitiva compuesta por 25 ítems, 10 ítems para el componente *Afectivo*, 8 para el *Cognitivo* y 7 para el *Comportamental*. (Estrada y cols. 2004, pág. 266).

Resultados

(a) Caracterización de la Muestra

La muestra total fue de 74 estudiantes, de la Licenciatura en Estadística fueron 36, (48.65%), de la Lic. en Informática fueron 38 (51.35%). La distribución de la muestra por grupos de sexo fue como sigue; en total fueron 27 hombres, (36.49%), y mujeres fueron 47 (63.51%). La distribución por grupos de sexo para cada programa de licenciatura fue como sigue; hombres, de la Lic. Estadística fueron 9 (12,16%) y mujeres fueron 27 (36.49%); de la Lic. en informática fueron 18 (24.32%), hombres y 20 (63.51%) mujeres. La edad de los participantes tuvo un valor mínimo de 18 y máximo de 27, con un promedio de 20.77 años, con una desviación estándar de 1.8. Las correspondientes estadísticas por grupos de sexo fueron muy semejantes...

(b) Confiabilidad del instrumento.

Se calculó el Alfa de Cronbach para fines de evaluar consistencia interna como indicador de la confiabilidad del mismo, obteniéndose un valor de .8025 para la escala total. Para la escala de AFECTIVO fue de 0.6825. Para la escala de COGNITIVO fue de .5809 Para la escala de CPORTAMETAL fue de .4670.

(c) Estadísticas descriptivas de los indicadores del instrumento.

Se obtuvieron los valores de las estadísticas media aritmética, median, desviación estándar. Los resultados fueron muy similares para los tres indicadores para los dos grupos determinados por las dos licenciaturas consideradas. (Ver tabla 1).

(d) Comparaciones por grupos de Programa

Se realizaron gráficos de perfiles distribucionales mediante gráficos de cajas y alambres para caracterizar las actitudes los grupos de las dos licenciatura por separado. En general los valores medios para las tres escalas resultan superiores al 60%. En particular el indicador de Cognitivo es el que puntúa más alto, le sigue AFECTIVO y COMPORTAMENTAL con valores más bajos. Esto fue así de manera muy semejante para los dos grupos de estudiantes de las dos licenciaturas consideradas. Los perfiles promedios de los dos grupos considerados muestran valores más altos para el grupo de estadística en los tres indicadores. (Ver figuras 1 a la 3, y tabla 2).

Las pruebas de hipótesis de diferencias de Medianas de Mann-Whitney y evidenciaron diferencia entre los dos grupos de licenciatura para los dos primeros componentes, no así para el tercero.

Conclusiones

De los resultados obtenidos podemos concluir que en este estudio y con base a la muestra utilizada, y el instrumento que se aplicó se encontró evidencia de que los dos grupos de licenciatura, Estadística e informática, muestran evidencia de tener diferentes perfiles grupales semejantes con nivel superior para el grupo de estadística, con diferencias significativas en los dos primeros indicadores AFECTIVO y COGNITIVO, pero sin diferencia en el COMPORTAMENTAL.

REFERENCIAS

- Mills, J. D. (2004). *College Student Journal*; Vol. 38 Issue 3, p349-361.
- Bazán, J. L. y Aparicio, A. S. (2006). Las actitudes hacia la Matemática-Estadística dentro de un modelo de aprendizaje. *Revista Semestral del Departamento de Educación*, Vol. XV N° 28, pp. 1-12.
- Estrada A., Batanero, C., y Fortuny, J. M., (2004). *Enseñanza de las Ciencias*, 22(2), 263 – 274.
- Haddock, G. & Maio, G. R. (Eds.). (2004). *Contemporary Perspectives on the Psychology of Attitudes: The Cardiff Symposium*. Hove, England: Psychology Press. Retrieved May 20, 2009, from Questia database: <http://www.questia.com/PM.qst?a=o&d=108881658>
- Mills, J. D. (2004). Students' attitudes toward statistics: implications for the future. *College Student Journal*; Vol. 38 Issue 3, p349-361.
- Oskamp, S., & Schultz, P. W. (2004). *Attitudes and Opinions* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates. Retrieved May 20, 2009.
- Ursini, S., Sánchez, G, Orendain, M., (2004). Validación y confiabilidad de una escala de actitudes hacia las matemáticas y hacia las matemáticas enseñadas con computadora. *Educación Matemática*. Vol. 8, núm. 3, pp. 59-78.

TABLAS Y FIGURAS

Tabla 1.- Estadísticas globales y por grupo de programa de licenciatura de los indicadores AFECTIVO, COGNITIVO, COMPORTAMENTAL.

| | | N | Media | Mínimo | Máximo | Dev. Estand. |
|-------------|--------|----|-------|--------|--------|--------------|
| GLOBAL | AFEC_P | 74 | 70,75 | 40,00 | 96,00 | 10,06 |
| | COG_P | 74 | 75,06 | 50,00 | 92,50 | 8,88 |
| | COMP_P | 74 | 66,52 | 45,71 | 88,57 | 9,97 |
| ESTADISTICA | AFEC_P | 36 | 74,05 | 46,00 | 96,00 | 9,95 |
| | COG_P | 36 | 77,29 | 50,00 | 92,50 | 9,60 |
| | COMP_P | 36 | 67,46 | 45,71 | 88,57 | 10,80 |
| INFORMATICA | AFEC_P | 38 | 67,63 | 40,00 | 84,00 | 9,23 |
| | COG_P | 38 | 72,96 | 52,50 | 87,50 | 7,68 |
| | COMP_P | 38 | 65,63 | 48,57 | 88,57 | 9,16 |

Tabla 2.- Pruebas de Mann-Whitney para diferencias de medianas entre grupos de licenciaturas para los tres indicadores.

| | N | N | Sum Rangos | Suma Ranos | | | |
|--------|---------|---------|------------|------------|--------|-------|---------|
| | Grupo 1 | Grupo 2 | Grupo 1 | Grupo 2 | U | Z | p-level |
| AFEC_P | 36 | 38 | 1611,00 | 1164,00 | 423,00 | 2,822 | ,0047 |
| COG_P | 36 | 38 | 1564,00 | 1211,00 | 470,00 | 2,314 | ,0206 |
| COMP_P | 36 | 38 | 1443,50 | 1331,50 | 590,50 | 1,011 | ,3119 |

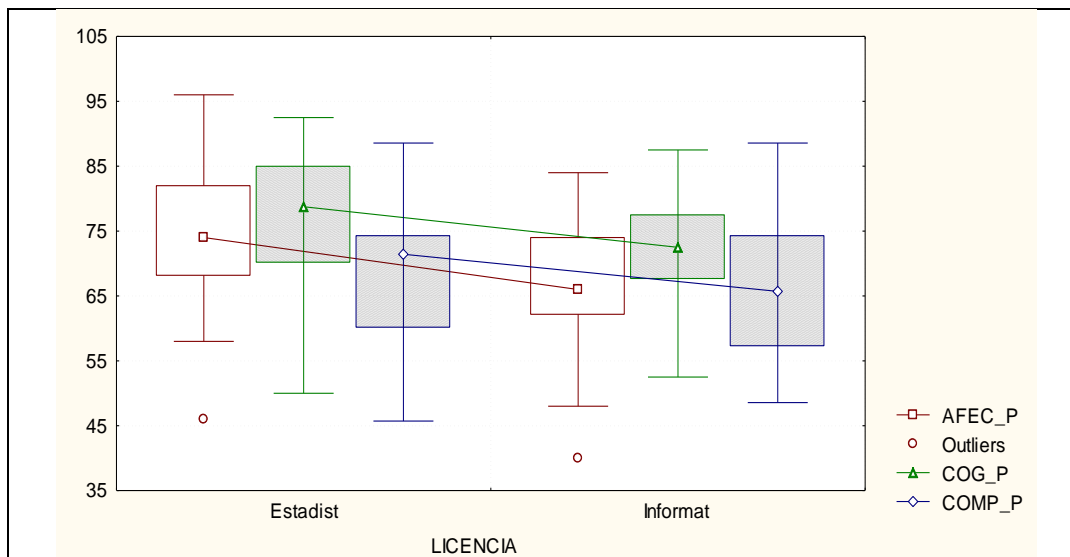


Figura 1- Perfiles distribucionales de los cuatro indicadores para el grupo de Estadística.

Propuesta de una Escala Alternativa para Medir Hábitos de Estudio en el Universitario

Ma. de los Ángeles Silva Mar¹, María de los Ángeles García Oramas², Claudio Rafael Castro López³

¹ Facultad de Pedagogía. Poza Rica. UV. asilva@uv.mx

² Facultad de Estadística e Informática. UV. angels1102@hotmail.com

³ Facultad de Estadística e Informática. UV. ccastro@uv.mx

Resumen

La Biometría nos brinda la oportunidad de aplicar métodos estadísticos y matemáticos en uno de los campos relacionados con la Psicología y la Pedagogía: El rendimiento de los estudiantes universitarios y sus hábitos de estudio.

Los Hábitos de estudio, según Díaz Vega 2000, en su propuesta para medir dichas condiciones considera las dimensiones siguientes: Distribución de tiempo (DT), Motivación en el estudio (ME), Actitudes y Conductas productivas en el estudio (AC) Preparación en los exámenes (PE), Notas de clase (NC), Distractores en el estudio (DE) y Optimización de la lectura (OP). Cada dimensión es medida con 10 ítems (el instrumento está diseñado en total con 70), que proporcionan un diagnóstico de condiciones de estudio identificados como: Adecuados o Inadecuados, según las respuestas emitidas por el entrevistado.

Se presenta lo que consideramos una importante contribución, al reducir considerablemente los ítems de la escala original, para clasificar estos hábitos de estudio, el cual tiene altos indicadores de confiabilidad y validez del instrumento, la reducción se ha realizado bajo un modelo de regresión logística, el cual tiene altos porcentajes de clasificación adecuada que impactan favorablemente en la bondad de ajuste del modelo. El ámbito de aplicación es de la Facultad de Pedagogía, campus Poza Rica-Tuxpan de la Universidad Veracruzana

1. Introducción

Al observar las dificultades mostradas en el comportamiento de los alumnos para lograr un desenvolvimiento acorde a las exigencias de la disciplina elegida y como futuros profesionistas surge la inquietud y preocupación de abordar la temática de abordar los hábitos de estudio que poseen los estudiantes. Diversas investigaciones han demostrado la carencia de metodologías para el estudio eficiente, con frecuencia los docentes solicitan a los estudiantes diversas tareas escolares y los resultados obtenidos en éstas, están relacionadas con los tipos de hábitos de estudio que poseen y que pueden ser identificados como hábitos de adecuados o inadecuados.

Una de las principales causas de los altos índices del fracaso académico en México es el inadecuado desarrollo de los hábitos de estudio desde el nivel educativo básico (Mackenzie, 1997).

Podemos señalar que los hábitos de estudio adecuados se les definen como la disposición adquirida para instrumentar estrategias, procedimientos y recursos encaminados a facilitar aprendizajes significativos, mientras que los hábitos inadecuados como la costumbre de implementar procedimientos y estrategias que impiden el logro de aprendizajes significativos.(Silva, 1996).

Díaz Vega en su libro *Aprende a estudiar con Éxito* realiza una propuesta de autoevaluación de sus condiciones de estudio y a partir del diagnóstico obtenido el estudiante puede conocer si sus condiciones de hábitos se ubican en adecuados o inadecuados y a partir de las condiciones identificadas propone unas sencillas guías que le garanticen al alumno un estudio eficaz. Pretenden medir estos hábitos mediante un conjunto de cuestionarios constituidos por preguntas que están agrupadas por áreas que van desde 3, 4 hasta 7 donde se va dando puntajes para cada una de las opciones de respuesta para todas y cada una de las acciones que constituyen a los hábitos de estudio, Díaz Vega propone un instrumento en el cual se tienen reactivos (preguntas) deseables o indeseables (*), este instrumento se divide en 7 áreas: *Distribución de tiempo* (DT), *Motivación en el estudio* (ME), *Actitudes y Conductas productivas en el estudio* (AC), *Preparación en los Exámenes* (PE), *Notas de Clase* (NC), *Distraedores en el Estudio* (DE) y *Optimización de la lectura* (OP), cada una integrada por 10 ítems que se responden en categorías: siempre, a menudo, raras veces y nunca, sus puntos van de acuerdo al tipo de reactivo, si es deseable o indeseable (*):

9. ¿Siento satisfacción al intervenir en actividades relacionadas con el estudio?

Siempre 3 A menudo 2 Raras veces 1 Nunca 0

3. ¿Hay personas conversando o ruidos que me molesten o distraigan mientras estudio?*

Siempre 0 A menudo 1 Raras veces 2 Nunca 3

De esta forma de acuerdo a la categoría que elija se asignara el valor correspondiente, al sumar los puntos de los 10 ítems por área, se obtiene el puntaje para conocer si el individuo tiene un hábito adecuado o inadecuado en esta área, siendo el criterio: *un puntaje menor o igual a 21 un hábito inadecuado y de 22 a 30 puntos adecuando*, así por cada una de las 7 áreas.

2. Problema

Tradicionalmente se ha considerado que el estudiante al ingresar a la Universidad debe poseer de manera sólida y sistemática formas metodológicas para su aprendizaje, así como cierta instrumentación de procedimientos que le faciliten su aprovechamiento escolar, esto fundamentalmente en la idea de que al haber tenido la experiencia a lo largo de varios años como estudiante se le ha presentado la necesidad de adoptar técnicas y métodos acordes a las exigencias académicas para optimizar al máximo su rendimiento escolar. Sin embargo, lo anterior no es del todo cierto, porque los alumnos muestran a lo largo de su vida universitaria una serie de problemas referidos al bajo aprovechamiento, reprobación e incluso deserción escolar.

De los recursos con que cuentan los estudiantes para el logro de los objetivos de estudios profesionales, el estudio es de los más importantes. El estudio es un recurso que nos permite apropiarnos de información, de aprender. El saber estudiar, -el hábito de la lectura- es una herramienta indispensable para todo estudiante universitario. Lamentablemente son pocos los que tienen un buen hábito de estudio, la adquisición de buenos hábitos deben redundar en poner también orden en nuestra vida, en una mejor organización de nuestra vida cotidiana, en una mayor productividad (De Gasperín, 1995).

Medir hábitos de estudio es complejo, por ejemplo a detección de las variables relevantes para conocer los hábitos de un estudiante (sus antecedentes familiares, de formación en educación básica y media), entre muchos otros. Es un factor importante.

3. Metodología

Un modelo logístico se utiliza para averiguar la probabilidad de ocurrencia de una (o más) de las dos categorías que tiene la variable respuesta o dependiente, en función de una o varias variables explicatorias o independientes (Silva, Barroso, 2004). La naturaleza de la variable respuesta en este caso dicotómica (hábitos de estudios adecuados o inadecuados) nos lleva a proponer el modelo logístico como la vía aplicada para este problema ya que sus características son:

- Los errores se distribuyen binomialmente
- No requiere de supuestos
- Cuenta con contrastes estadísticos directos
- Capacidad de incorporar efectos no lineales
- Amplia variedad de diagnóstico

Se obtienen diversos modelos en función de las variables independientes (los 70 ítems de los hábitos de estudio) para ubicar a un individuo en alguna de estas dos categorías de la variable respuesta, en este caso lo que se pretende es detectar aquellas variables que contribuyen de manera determinante para clasificar los hábitos de un estudiante como adecuados e inadecuados. El modelo general es:

$$y_i = \frac{1}{1 + e^{-\beta_1 + \beta_2 x_i}}$$

Donde:

y_i : variable respuesta, se tiene un valor de corte para cada categoría de 0.5 arriba o 0.5 abajo

β_i : Estimadores

e : base de logaritmos naturales

Para el caso de nuestra investigación, un ejemplo de modelo inicial es:

$$DT_i = p1+p8+p15+p22+p29+p36+p43+p50+p57+p64$$

De esta forma para clasificar los hábitos de un individuo, en el área de DT (distribución de tiempo) el modelo aditivo incorpora cada uno de los reactivos de acuerdo al criterio de aquellos que son deseables o indeseables (*); en base a una muestra de 112 estudiantes de la Facultad de Pedagogía campus Poza Rica-Tuxpan de la Universidad Veracruzana, se aplicó un modelo de regresión logística, *teniendo como resultado una probabilidad de ocurrencia de cada uno de los eventos*, con un punto de corte de 0.5; una vez incorporados los modelos en el programa SPSS v11.0 y obteniendo las tablas de valoración correspondientes, del modulo de regresión logística utilizando el Método de paso hacia delante Wald (Visauta,1998) como resultado se obtienen los ítems significativos para clasificar los hábitos de un individuo, la Tabla 1 muestra estos ítems por área:

Tabla 1. Preguntas originales del cuestionario vs preguntas propuestas por el modelo de regresión logística.

| DT | | ME | | DE | | NC | | OP | | PE | | AC | |
|-------|------------|-------|------------|-------|------------|-------|------------|-------|------------|-------|------------|-------|------------|
| Orig. | Mod. Prop. | Orig. | Mod. Prop. | Orig. | Mod. Prop. | Orig. | Mod. Prop. | Orig. | Mod. Prop. | Orig. | Mod. Prop. | Orig. | Mod. Prop. |
| 1 | 1 | 2 | 51 | 3* | 3* | 4 | 11 | 5 | 26 | 6 | 13 | 7* | 7* |
| 8 | 8 | 9 | | 10* | 24* | 11 | 18 | 12 | 40 | 13 | 20 | 14* | 14* |
| 15 | 15 | 16* | | 17* | 52 | 18 | 25 | 19 | 61 | 20 | 27 | 21 | 28 |
| 22 | 22 | 23* | | 24* | | 25 | 39 | 26 | 68 | 27 | 62 | 28 | 56 |
| 29 | 64 | 30 | | 31* | | 32 | 60 | 33* | | 34* | 69 | 35 | 63* |
| 36* | | 37* | | 38* | | 39 | 67 | 40 | | 41 | | 42 | |
| 43 | | 44* | | 45 | | 46 | | 47* | | 48 | | 49 | |
| 50 | | 51 | | 52 | | 53 | | 54 | | 55 | | 56 | |
| 57 | | 58 | | 59* | | 60 | | 61 | | 62 | | 63* | |
| 64 | | 65* | | 66 | | 67 | | 68 | | 69 | | 70 | |

Por ejemplo, para diagnosticar los hábitos de estudio de un individuo en el área Distribución de tiempo, se tiene:

$$DT_i = \frac{1}{1 + e^{-(1146.43 - 152.40(p_1) - 83.83(p_8) - 129.74(p_{15}) - 69.46(p_{22}) - 92.09(p_{64}))}}$$

Las variables que contribuyen de manera significativa en el modelo anterior son los ítems 1, 8, 15, 22 y 64; en la Tabla 2 se encuentran las 7 áreas con sus respectivos modelos:

TABLA 2. Modelos de las 7 áreas

$$DT_i = \frac{1}{1 + e^{-(1146.43 - 152.40(p1) - 83.83(p8) - 129.74(p13) - 69.46(p22) - 92.09(p64))}}$$

$$ME_i = \frac{1}{1 + e^{-(32.89 - 10.27(p31))}}$$

$$DE_i = \frac{1}{1 + e^{-(19.28 - 2.47(p3) - 2.74(p24) - 3.02(p52))}}$$

$$[NC] \quad i = 1 / (1 + e^{(-(-37.11 + 2.22(p11) + 3.01(p18) + 2.47(p25) + 3.10(p39) + 3.60(p60) +$$

$$OP_i = \frac{1}{1 + e^{-(18.83 - 2.68(p26) - 1.99(p40) - 1.90(p61) - 2.22(p68))}}$$

$$PE_i = \frac{1}{1 + e^{-(23.30 - 2.10(p13) - 1.43(p20) - 2.49(p27) - 2.43(p62) - 1.44(p69))}}$$

$$AC_i = \frac{1}{1 + e^{-(33.05 - 1.94(p7) - 2.46(p14) - 4.86(p28) - 2.62(p56) - 1.27(p63))}}$$

Con estos modelos se obtuvieron los siguientes valores de ajuste para cada área:

| Área | Bondad de Ajuste (%) | Área | Bondad de Ajuste (%) |
|------|----------------------|------|----------------------|
| DT | 99.1 | OP | 90.2 |
| ME | 96.4 | PE | 87.5 |
| DE | 95.5 | AC | 95.5 |
| NC | 92.9 | | |

A partir de estos alentadores resultados, como una segunda etapa se desarrolla un proyecto con el claustro ya mencionado, que consiste en utilizar una plataforma en línea permitiendo que el cuestionario de hábitos de estudio fuese contestado en línea, haciendo uso de un sistema masivo que permitió captar un muestra de 358 alumnos a los cuales se ha aplicado el modelo, obteniendo los siguientes valores de bondad de ajuste:

| Área | Ajuste Total (%) | Error Total (%) | Área | Ajuste Total (%) | Error Total (%) |
|------|------------------|-----------------|------|------------------|-----------------|
| DT | 97.5 | 2.5 | OP | 93.9 | 6.1 |
| ME | 98.6 | 1.4 | PE | 100 | 0 |
| DE | 93.6 | 6.4 | AC | 100 | 0 |
| NC | 100 | 0 | | | |

4. Conclusiones

Ante la complejidad de la medición de hábitos de estudio, se ha obtenido una propuesta metodológica para su medición en un estudiante universitario, reduciendo sustancialmente la propuesta de Díaz Vega (70 ítems) , reduciendo considerablemente la escala (29 ítems) en base a un modelo logístico. Se han configurado las etapas iniciales para presentar una propuesta institucional que mejore la detección de los buenos hábitos de estudio en un estudiante universitario.

REFERENCIAS

- Beltrán Casanova J., Suárez Domínguez José L.(2003); El quehacer tutorial, Talleres Integra Tecnologías Gráficas; Xalapa, Veracruz.
- Castro Posada J.A. (1999), Técnicas de estudio para universitarios un reto para tu autoformación, Amarú ediciones, Salamanca.
- De Gasperín Roberto (1995) “Manual de hábitos de estudio para estudiantes (no para alumnos) Texto Universitario
- Díaz Vega José Luis (2000), Aprende a estudiar con éxito, Edit. Trillas, México, D.F.
- Visauta Vinacua B. (1998), Análisis estadístico con SPSS para Windows, Edit. Mc Graw Hill, España.
- Silva A. Luis C., Barroso Utra Isabel M. (2004), Regresión Logística, Cuadernos de Estadística No. 27, Edit. La Muralla, S.A., Madrid.
- Silva Mar M. A. (1996); “Los Hábitos de estudio con que ingresan los estudiantes de Pedagogía de la Universidad Veracruzana y su influencia en la Formación Académica” Tesis para Obtener Grado de Maestría, Tampico, Tamps.

Interpretación y Construcción de Gráficas: Dificultades del Estudiante Egresado del Bachillerato

Abraham Cuesta Borges¹, Marco Antonio Méndez Salazar²

Facultad de Economía, Universidad Veracruzana

¹acuesta@uv.mx

²marcomendez@uv.mx

Resumen

El presente trabajo muestra los resultados de una investigación (Cuesta, 2007), cuyos sujetos de estudio fueron estudiantes de la Licenciatura en Economía de la Universidad Veracruzana, con el objetivo de conocer las dificultades en la comprensión del concepto de función, dentro del tema funciones y gráficas. Es un estudio diagnóstico realizado a estudiantes que deben acreditar la asignatura Cálculo I (el curso de cálculo diferencial e integral con funciones de una variable), cuyo objetivo fue conocer qué conocimientos previos pueden causar dificultades en el proceso de aprendizaje del concepto función. Se propone a los estudiantes una serie de situaciones relacionadas con el entorno gráfico y representando la dependencia funcional entre variables. Los resultados muestran la existencia de dificultades en la interpretación y construcción de gráficas contextualizadas, muchas de ellas provocadas por un sistema de enseñanza que se desarrolla a través de las exposiciones del profesor, seguido del intento por resolver problemas y ejemplos rutinarios del libro de texto.

1. Introducción

La problemática referente al proceso de aprendizaje de los conceptos del cálculo diferencial e integral plantea múltiples interrogantes a la investigación en didáctica de las matemáticas. En el caso especial de la enseñanza universitaria, uno de los conceptos más importantes es, sin duda, el de función; de su comprensión depende que se adquieran competencias para el modelado de situaciones directamente relacionadas con el campo de estudio y para plantear y resolver problemas de optimización. Es precisamente este concepto el foco de la investigación que, basada en la línea del Pensamiento Matemático Avanzado (PMA), pone especial énfasis al estudio de las construcciones y los procesos mentales que, sustentados en la experiencia previa del estudiante, afectan su aprendizaje significativo. Se entiende aquí por experiencia previa tanto aquellas situaciones que no están vinculadas con las construcciones matemáticas, como aquellas que se sustentan en los conocimientos matemáticos que adquiere el estudiante dentro de las instituciones educativas.

Los resultados de la investigación (Cuesta, 2007) muestran la existencia de dificultades en la interpretación y construcción de gráficas. Se ponen de manifiesto dos confusiones, planteadas por Janvier (1987a): la confusión gráfico-dibujo y la confusión verbal-gráfico. Los estudiantes no reconocen la dependencia funcional entre las variables implicadas en la gráfica; por otra parte, existe la tendencia a asumir linealidad, pero se pierde el significado de la pendiente de una recta. Estas dificultades, también examinadas en Leinhardt,

Zaslavsky y Stein (1990) y Deulofeu (1993, 1995) son producidas por el efecto combinado de los significados que poseen los estudiantes sobre el concepto de función y del conocimiento que tienen sobre los contextos en que se deben realizar dichas tareas.

2. Referencias de investigación

Los estudiantes ingresan a la Universidad después de culminados sus estudios de secundaria y bachillerato, cuyos planes de estudios oficiales son establecidos por la Secretaría de Educación Pública (SEP) de México. De acuerdo con el programa de estudio oficial para la asignatura de Matemáticas IV de bachillerato, la enseñanza tiene como uno de sus propósitos formativos: “el desarrollo de conocimientos, habilidades y actitudes que le permitan interpretar de manera reflexiva y crítica el quehacer científico, su importancia actual y futura [...] al que contribuye el estudio de esta asignatura mediante el desarrollo de conceptos, métodos y procesos lógicos que permiten al estudiante analizar la relación funcional entre dos variables, indispensable para la explicación de fenómenos y la resolución de problemas en distintos campos del conocimiento” (SEP, 2005). En la práctica, sin embargo, se exige que el alumno adquiriera conocimientos, casi siempre por transferencia directa de un maestro, seguida del intento de los estudiantes por resolver problemas y ejemplos rutinarios del libro de texto. Como resultado, para la mayoría de los estudiantes la forma más segura de enfrentarse con éxito al dominio matemático no es intentar comprender, sino simplemente comportarse mecánicamente; las ideas permanecen borrosas, incoherentes y poco adaptadas a las necesidades del mundo del cálculo.

En contraste con esta realidad, las investigaciones actuales en didáctica de las matemáticas (Artigue, 1995), ponen especial énfasis al estudio del proceso cognitivo, así como a las diferentes representaciones, físicas o mentales, que son imprescindibles y dan significado a los conceptos matemáticos estudiados. En este sentido, Dreyfus (1991) señala que el proceso de aprendizaje se realiza en etapas; por ejemplo, el concepto de función, en cuyo aprendizaje (Azcárate y Deulofeu, 1990) se hallan varias representaciones (gráficas, expresiones algebraicas, tablas, diagramas de flechas, etc.). Precisamente, en los trabajos de Janvier (1978, 1987a, 1987b) se abordan las diferentes traducciones entre distintos tipos de representaciones externas del concepto matemático (verbal, tabla, gráfica y expresión algebraica), así como las representaciones internas del estudiante, es decir, la combinación de tres componentes: símbolos escritos, objetos reales e imágenes mentales, que interactúan con las representaciones externas del conocimiento (Dreyfus, 1991).

Se toman como referencia las aportaciones de algunos investigadores en la línea PMA: Tall y Vinner (1981), sobre la formación de conceptos; Leinhardt et al. (1990), donde se analiza la naturaleza del aprendizaje en términos de intuiciones y errores conceptuales de los estudiantes; Azcárate (1992), respecto a los esquemas conceptuales de los alumnos en relación con el concepto de pendiente de una recta; Deulofeu (1993), donde se relacionan las tareas de interpretación y construcción con el lenguaje gráfico y Artigue (1995), sobre las diferentes representaciones, físicas o mentales, que dan significado a los conceptos matemáticos.

3. Metodología

La metodología seguida en el estudio fue cualitativa, de naturaleza empírica y exploratoria. Se propuso a 58 estudiantes un conjunto de situaciones planteadas en una prueba escrita, con el objeto de conocer y analizar las dificultades en tareas de interpretación y/o construcción de gráficas. De este grupo se extrajo una muestra de seis estudiantes para el estudio de casos, conducido a través de entrevistas en sesiones grupales, con la finalidad de conocer e interpretar las causas de dichas dificultades en un contexto de intercambio de ideas, que permitió obtener datos sobre la forma en que los estudiantes llegaron, individual y colectivamente, a la solución de los problemas planteados. A continuación se muestran dos de las situaciones planteadas a los estudiantes en la prueba escrita:

Situación 1 (tarea de interpretación): La Figura 1 muestra las gráficas “distancia contra tiempo” resultantes de la observación del movimiento de dos coches, el A y el B.

- (a) Se desea saber si, en el instante de tiempo $t = 3$, la velocidad del coche A es menor, igual o mayor que la velocidad del coche B. Explica.
- (b) ¿Qué ocurre con las velocidades de los coches en el instante de tiempo $t = 6$? ¿Cuál es mayor? ¿Son iguales? Explica.

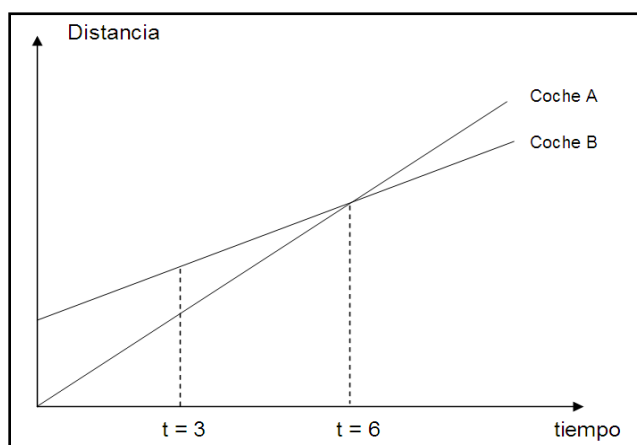


Figura 1

Situación 2 (tarea de construcción): La Figura 2 muestra el recorrido en una montaña rusa. La pista por donde viaja el carrito está señalada con diferentes puntos: A, B, C, D, E, F y G. Se sabe que entre los puntos A y B el carrito viaja a una velocidad lenta y constante.

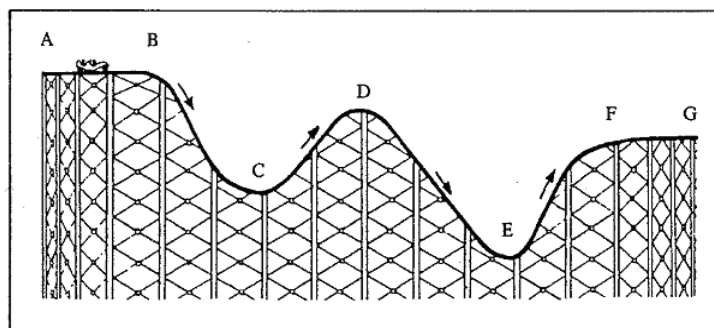


Figura 2

- (a) ¿Cómo varía la velocidad del carrito durante el recorrido que va desde el punto A hasta el punto G? Explica por escrito, con tus palabras.
- (b) Describe tu respuesta anterior mediante una gráfica.

4. Resultados

Las respuestas aportan una idea respecto a las dificultades en tareas relacionadas con la interpretación y construcción de gráficas que relacionan dos magnitudes, en contextos supuestamente conocidos por el estudiante. En las respuestas al inciso (a) de la Situación 1, se pudo observar que el 40% de los estudiantes no reconoce la velocidad como una relación entre las magnitudes distancia y tiempo, cuya relación se explica a través del concepto de pendiente de una recta. En contraste, se relaciona la velocidad con la altura medida sobre el eje vertical: en opinión del 80% de los estudiantes, el coche A tiene menor velocidad que el coche B en el instante de tiempo $t = 3$. Algunas respuestas (erróneas) son: “B ya tiene velocidad, no parte del origen como A, está arriba en el eje”, “B estaba situado más lejos y llevaba mayor velocidad”, “B no parte del origen, recorrió mas distancia en el mismo tiempo”.

En respuesta al inciso (b) de la Situación 1, se plantea que las velocidades de ambos coches se igualan en el instante $t = 6$; en este caso, la imagen visual de dos líneas que se unen imposibilita una interpretación adecuada de la velocidad. Algunos argumentos son: “Ambos coches tienen la misma distancia y tiempo”, “los coches se encuentran en el mismo lugar”, “los dos coches llegan al mismo punto, se encontraron en $t = 6$ ”, “al tiempo $t = 6$ los dos coches se interceptan”.

Una situación similar se manifiesta en las respuestas a la Situación 2. El 28% de los estudiantes realiza una interpretación icónica de los cambios de la velocidad del coche; la grafica proporcionada por ellos es una representación (un dibujo) de la montaña rusa que, esencialmente, copia la Figura 2. Otros estudiantes representan de manera correcta los cambios de velocidad, pero se utiliza un prototipo al asumir un modelo lineal (trazando segmentos de línea recta) para representar las diferentes fases del recorrido.

5. Conclusiones

Los resultados muestran que, existen dificultades en tareas de interpretación y construcción, producidas por el efecto combinado de los significados que poseen los estudiantes sobre este concepto y del conocimiento que se tiene sobre los contextos en que se deben realizar dichas tareas. Las dificultades más relevantes son:

1. La incomprensión de la relación entre variables mostradas en la gráfica: La situación contextualizada se analiza atendiendo solamente a una representación literal de las magnitudes involucradas, sin establecer la relación manifiesta entre ellas.
2. La incomprensión del significado de la pendiente de la recta: No se reconoce, en la gráfica de una situación contextualizada, la velocidad como la pendiente de la recta; identificándose dicha variable con la altura vertical de la recta en la gráfica.

3. La interpretación icónica de situaciones gráficas contextualizadas: Muchos estudiantes realizan una interpretación icónica de los cambios de la velocidad de un coche que sube y baja una montaña rusa; para estos estudiantes la gráfica que debe representar los cambios de velocidad es, en esencia, una representación (una imagen) de la montaña rusa.

REFERENCIAS

- Artigue, M. (1995). La enseñanza de los principios del cálculo: Problemas epistemológicos, cognitivos y didácticos. En: *Ingeniería didáctica en educación matemática*, M. Artigue, R. Douady, L. Moreno y P. Gómez (eds.), 97-140. Bogotá: Grupo Editorial Iberoamérica.
- Azcárate, C. (1992). Estudio de los esquemas conceptuales y de los perfiles de unos alumnos de segundo de BUP en relación con el concepto de pendiente de una recta. *Epsilon*, **24**, 9-22.
- Azcárate, C. y Deulofeu, J. (1990). *Funciones y gráficas*. Madrid: Editorial Síntesis.
- Cuesta, A. (2007). *El proceso de aprendizaje de los conceptos de función y extremo de una función en estudiantes de economía: análisis de una innovación educativa*. Tesis doctoral no publicada. Bellaterra: Universidad Autónoma de Barcelona.
- Deulofeu, J. (1993). *Els Gràfics cartesianes de funcions: un estudi de les concepcions dels alumnes centrat en el significat del gràfic*. Tesis doctoral no publicada. Bellaterra: Universidad Autónoma de Barcelona.
- Deulofeu, J. (1995). Concepciones de los alumnos de secundaria sobre distintas gráficas de funciones. *UNO. Revista de Didáctica de las Matemáticas*, **4**, 6-16.
- Dreyfus, T. (1991) Advanced mathematical thinking processes. In: *Advanced Mathematical Thinking*, D. Tall (ed.), 25-41. Dordrecht: Kluwer, A. P.
- Janvier, C. (1978). *The interpretation of complex cartesian graphs representing situations*. Unpublished Ph. D. Thesis. University of Nottingham.
- Janvier, C. (1987a). Representations and understanding: The notion of function as an example. In: *Problems of representation in the teaching and learning of mathematics*, C. Janvier (ed.), 67-72. NJ: Lawrence Erlbaum Associates.
- Janvier, C. (1987b). Translation processes in mathematics education. In: *Problems of representation in the teaching and learning of mathematics*, C. Janvier (ed.), 27-32. NJ: Lawrence Erlbaum Associates.
- Leinhardt, G., Zaslavsky, O. & Stein, M. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, **60**(1), 1-64.
- Secretaría de Educación Pública, Subsecretaría de Educación Media Superior, Dirección General de Bachillerato (2005). *Programa de estudio de Matemáticas IV*. Obtenido el 18 de marzo de 2009 desde http://www.dgb.sep.gob.mx/informacion_academica/programasdeestudio/cfb_4osem/Matematicas-IV.pdf
- Tall, D. & Vinner, S. (1981). Concept image and concept definition in mathematics, with special reference to limits and continuity. *Educational Studies in Mathematics*, **12**, 151-169.

Entornos Geométrico y Algebraico: Dificultades del Estudiante Egresado del Bachillerato

Abraham Cuesta Borges¹, Marco Antonio Méndez Salazar²

Facultad de Economía, Universidad Veracruzana

¹acuesta@uv.mx

²marcomendez@uv.mx

Resumen

El presente trabajo muestra los resultados de investigaciones (Cuesta 2005, 2007) realizadas con el objetivo de conocer los obstáculos en la comprensión de la matemática universitaria. Dentro del tema funciones y gráficas, se propone a los estudiantes que deben cursar la asignatura Cálculo I (cálculo diferencial e integral de funciones de una variable) un conjunto de situaciones, en los contextos geométrico y algebraico, relacionadas con las diferentes formas de representar el concepto de función. Los estudios dan constancia sobre la existencia de dificultades en el proceso de aprendizaje, muchas de ellas provocadas por el propio sistema de enseñanza previo a la universidad.

1. Introducción

En el ámbito de la enseñanza universitaria uno de los conceptos más importantes es, sin duda, el de función; de su comprensión depende que se adquieran competencias para plantear y resolver problemas elementales de optimización. Nuestras investigaciones reportan algunos resultados sobre las dificultades del estudiante en la comprensión del concepto de función; en particular aquellas relacionadas con la incompreensión de la dependencia funcional entre variables. Se propuso, a estudiantes que iniciaban la Licenciatura en Economía en la Universidad Veracruzana, un conjunto de situaciones contextualizadas y relacionadas con las diferentes formas de representar el concepto de función, con el propósito de conocer y analizar las dificultades, en el proceso de aprendizaje, a partir de tareas relacionadas con su interpretación y construcción.

La investigación didáctica, enfocada a los problemas en el proceso de aprendizaje y en línea del Pensamiento Matemático Avanzado (PMA), centra especial atención al estudio de las construcciones y los procesos mentales que, sustentados en la experiencia previa del estudiante, afectan el aprendizaje significativo de nuevos conceptos matemáticos. Entendiendo por experiencia previa tanto aquellas que no están vinculadas con las construcciones matemáticas, como aquellas experiencias que se sustentan en los conocimientos matemáticos que adquiere el estudiante en la institución educativa.

Las conclusiones ponen de manifiesto que el esquema conceptual (Tall y Vinner, 1981) no permite acceder al estudio de las funciones. Los estudiantes egresados del bachillerato no pueden entender, tal y como exponen Leinhardt, Zaslavsky y Stein (1990), una definición a partir de la otra y no entienden la idea intuitiva (usando el término ‘intuitivo’ como antítesis de ‘riguroso’) básica del concepto de función; es decir, que variar los valores de una

variable (independiente) provoca un cambio en los valores de una segunda variable (dependiente).

2. Referencias Teóricas

La investigación se desarrolla en el marco de las aportaciones del PMA, con especial atención a los trabajos de Tall y Vinner (1981) sobre la formación de conceptos (“esquema conceptual” y “definición conceptual”), entendiendo que el esquema conceptual es “la estructura cognitiva de un individuo, asociada a un concepto matemático, y se define como el conjunto de todas las imágenes mentales (cualquier clase de representación: forma simbólica, diagrama, gráfica, etc.) del estudiante asociadas al concepto con todas las propiedades y procedimientos que le caracterizan” (Azcárate, 1995, p. 55). Cuyas imágenes, de acuerdo con Artigue (1990) están sustentadas en la experiencia del estudiante, poseedor de una concepción personal caracterizada por: (i) la clase de situaciones-problemas que dan sentido al concepto, (ii) el conjunto de significados que el estudiante es capaz de asociarle, y (iii) los instrumentos, teoremas, y algoritmos de los que dispone el estudiante para manejar el concepto.

También considera los trabajos de Janvier (1978, 1987), sobre las representaciones internas del estudiante, es decir, la combinación de tres componentes: símbolos escritos, objetos reales e imágenes mentales que interactúan con las representaciones externas del conocimiento (Dreyfus, 1991). La investigación en el campo de la didáctica de la matemática documenta la existencia de una amplia variedad de obstáculos en el aprendizaje de conceptos asociados con el estudio de funciones. En Artigue (1995), por ejemplo, se expone sobre dificultades en: (i) el manejo de conceptos abstractos, (ii) la comprensión de conceptos, y (iii) la interpretación de resultados. Existen otras referencias sobre obstáculos y dificultades en el proceso de aprendizaje, por ejemplo: Janvier (1987), Leinhardt et al. (1990), Azcárate (1992), Deulofeu (1993) y Lither (2003). En éstas se presupone que el interrogante acerca de cuáles son las dificultades en el aprendizaje del concepto de función es relevante en tanto permita desarrollar estrategias para la comprensión del significado del concepto a partir del conocimiento previo del sujeto.

3. Metodología

La metodología utilizada es cualitativa, de naturaleza empírica y exploratoria; basada en un experimento realizado con alumnos que inician sus estudios universitarios. Los datos se obtienen de las respuestas a un conjunto de tareas contextualizadas relacionadas con el concepto de función. Se dispuso de dos fuentes de obtención de datos: (i) las respuestas escritas de un grupo compuesto por 58 estudiantes, quienes responden a las preguntas planteadas en una prueba escrita, y (ii) el estudio de casos (muestra de seis estudiantes) con el objetivo de describir e interpretar la actuación, en trabajo cooperativo, cuando intercambian ideas y analizan la manera de abordar las situaciones propuestas. Para efectos de esta exposición, a continuación se transcriben dos de las preguntas planteadas a los alumnos, expresadas de manera verbal y geométrica:

Pregunta 1: La suma de los lados adyacentes de un rectángulo es 15 cm. Se desea estudiar cómo varía el área del rectángulo cuando varía la longitud de los lados.

- Realiza una tabla de valores, con los valores del lado frontal y los valores del área que se obtengan.
- ¿Cómo se puede representar esta relación mediante una gráfica?
- ¿Es esta relación una función? ¿Cuáles son las variables?
- Escribe una ecuación que permita hallar el valor numérico de la variable dependiente a partir del valor de la variable independiente.

Pregunta 2: El departamento de carreteras del gobierno municipal está planeando construir un área de camping al lado de una carretera principal. Debe ser rectangular con un área de 5,000 metros cuadrados y debe ser adyacente a la carretera, como se muestra en el dibujo (Figura 1 en este texto).

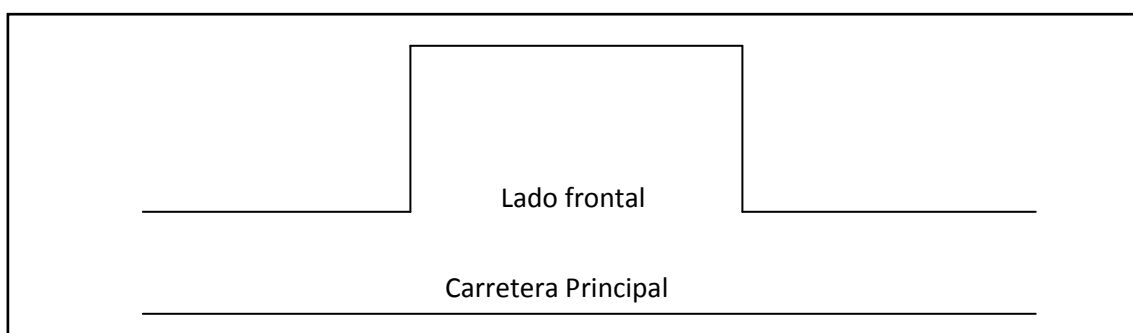


Figura 1: Dibujo a que se refiere la Pregunta 2.

Expresa el número de metros de cerca requerido como una función de la longitud del lado frontal. Para ello, lleva a cabo, una a una, las siguientes tareas:

- Construye una tabla de valores del número de metros de cercado, a partir de los valores que le asignes al lado frontal.
- Diseña un gráfico, a partir de los valores antes obtenidos, indicando la variable que colocadas en cada eje.
- Escribe una ecuación que exprese la relación entre la longitud del lado frontal a la carretera y el número de metros de cercado.
- ¿Cuál es la mínima longitud de cerca necesaria para cercar el camping?

4. Resultados

De inicio es importante analizar que los estudiantes, por lo general, no reconocen la existencia de una relación de dependencia entre las variables implicadas en situaciones contextualizadas. Se manifiesta una incomprensión del lenguaje matemático elemental; muchos estudiantes no comprenden la condición de la Pregunta 1: “la suma de dos lados adyacentes de un rectángulo es 15 cm”. Existe un obstáculo en la tarea de trasladar la idea geométrica al lenguaje de funciones. Se pudo observar que, de un total de 58 respuestas: (i) en 17 casos se representa la función, y (ii) de éstas, sólo en 7 respuestas se escribe una ecuación algebraica de la forma $A = x(15 - x)$, donde A representa el área del rectángulo y x la longitud del lado frontal. El resto de los estudiantes sólo pueden responder con una

ecuación en la forma $A = L \cdot L$ (“área igual a lado por lado”) o $A = b \cdot h$ (“área igual a base por altura”), hecho que resulta de un nivel de conocimiento muy elemental sobre los contextos geométrico y algebraico, que fueron objeto de estudios en secundaria y bachillerato.

En las respuestas a la Pregunta 2 se constata la imposibilidad de trasladar las ideas, expresadas en lenguaje común y contexto geométrico, al lenguaje de funciones. No se comprende la orden “expresa el número de metros de cerca requerido como una función de la longitud del lado frontal”, lo cual impide al estudiante ejecutar los procesos de pensamiento necesarios para concebir la relación funcional implícita. Otra dificultad surge del contexto geométrico: no se comprende la relación existente entre el área y el perímetro. Por esta razón, no se entiende el proceso de asignar valores al lado frontal, dado un valor del área, para determinar el lado lateral. Sin embargo, cuando se plantea la pregunta “¿Cómo se calcula el área?”, la respuesta de los estudiantes invariablemente es: “base por altura”. La memorización de fórmulas sin contexto conduce a una pérdida del sentido de la relación existente entre el perímetro y el área de un rectángulo.

En adición, existe incompreensión del lenguaje algebraico a partir del contexto geométrico planteado. En efecto, después de lograda la tabla y la representación gráfica, no se comprende el proceso para hallar la longitud de un lado del rectángulo, dados el área y la longitud del otro lado. Por esta situación no se puede responder a la orden “escribe una ecuación que exprese la relación entre la longitud del lado frontal a la carretera y el número de metros de cercado”.

5. Conclusiones

Los resultados muestran que, para el 80% de los estudiantes, resulta una dificultad cognitiva asociar el concepto de función a situaciones contextualizadas. La experiencia personal se basa, casi siempre, en lo que se estudia de manera inmediata y no se tiene la capacidad de utilizar el concepto, de una manera flexible, en situaciones conocidas o problemas que formen parte de las experiencias personales y conocimientos anteriores. El esquema conceptual (Tall y Vinner, 1981) evoca un aspecto del concepto: la existencia de una relación de dependencia entre dos variables, pero la idea que se tiene sobre la preimagen e imagen de la función causa confusión cuando se tiene que decidir, en un contexto concreto y supuestamente conocido, si la relación entre dos variables es una función. La falta de experiencia, sobre el contexto geométrico confirma la idea de Artigue (1990) respecto a que ciertas dificultades tienen una estrecha relación con la propia experiencia personal del estudiante.

Existen, además, dificultades en tareas de interpretación y construcción, producidas por el efecto combinado de los significados que poseen los estudiantes sobre este concepto y del conocimiento que se tiene sobre los contextos en que se deben realizar dichas tareas. La tarea de traducción, en el sentido de Janvier (1987), queda afectada cuando se desarrolla en un contexto geométrico. Se constata que el nivel de conocimiento sobre el contexto geométrico es muy elemental y crea un conflicto para construir las diferentes formas de representación del concepto de función, como el hecho de que muchos estudiantes no pueden transferir la función, de la tabla y/o gráfica, a la ecuación algebraica.

REFERENCIAS

- Artigue, M. (1990). Épistémologie et didactique. *Recherches en Didactique des Mathématiques* **10**(2,3), 241-286.
- Artigue, M. (1995). La enseñanza de los principios del cálculo: Problemas epistemológicos, cognitivos y didácticos. En: *Ingeniería didáctica en educación matemática*, M. Artigue, R. Douady, L. Moreno y P. Gómez (eds.), 97-140. Bogotá: Grupo Editorial Iberoamérica.
- Azcárate, C. (1992). Estudio de los esquemas conceptuales y de los perfiles de unos alumnos de segundo de BUP en relación con el concepto de pendiente de una recta. *Epsilon*, **24**, 9-22.
- Azcárate, C. (1995). Sistemas de representación. *Uno, Revista de Didáctica de las Matemáticas*, **4**, 13-20.
- Cuesta, A. (2005). *Dificultades de los estudiantes de economía en el aprendizaje del concepto de extremo de una función*. Tesis de maestría no publicada. Bellaterra: Universidad Autónoma de Barcelona.
- Cuesta, A. (2007). *El proceso de aprendizaje de los conceptos de función y extremo de una función en estudiantes de economía: análisis de una innovación educativa*. Tesis doctoral no publicada. Bellaterra: Universidad Autónoma de Barcelona.
- Deulofeu, J. (1993). *Els Gràfics cartesianes de funcions: un estudi de les concepcions dels alumnes centrat en el significat del gràfic*. Tesis doctoral no publicada. Bellaterra: Universidad Autónoma de Barcelona.
- Dreyfus, T. (1991) Advanced mathematical thinking processes. In: *Advanced Mathematical Thinking*, D. Tall (ed.), 25-41. Dordrecht: Kluwer, A. P.
- Janvier, C. (1978). *The interpretation of complex cartesian graphs representing situations*. Unpublished Ph. D. Thesis. University of Nottingham.
- Janvier, C. (1987). Representation and understanding: The notion of function as an example. In: *Problems of representation in the teaching and learning of mathematics*, C. Janvier (ed.), 67-72. Hillsdale, N. J.: Lawrence Erlbaum.
- Leinhardt, G., Zaslavsky, O. & Stein, M. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, **60**(1), 1-64.
- Lithner, J. (2003). Students' Mathematical Reasoning in University Textbook Exercises. *Educational Studies in Mathematics*, **52**(1), 29-55.
- Tall, D. & Vinner, S. (1981). Concept image and concept definition in mathematics, with special reference to limits and continuity. *Educational Studies in Mathematics*, **12**, 151-169.

Ocurrencia Temporal de Parásitos en *Macrobrachium acanthurus* de la Cuenca Baja del Río Jamapa

Magda Estela Domínguez-Machín¹, Sergio Francisco Juárez Cerrillo²

¹Facultad de Medicina Veterinaria y Zootecnia, U.V. magda-dominguez@hotmail.com

²Facultad de Estadística e Informática. U.V. sejuarez@uv.mx

Resumen

Se presentan los resultados del diagnóstico parasitológico en poblaciones silvestres del langostino *Macrobrachium acanthurus* provenientes de la cuenca baja del río Jamapa en Boca del Río (México), durante octubre del 2007 a julio del 2008. Los resultados indican la presencia de cinco especies de protozoarios parásitos: *Epistylis* sp., *Acineta* sp., *Lagenophrys* sp., gregarinas y un ciliado. Estos organismos se consideran epibiontes a excepción de las gregarinas, ya que habitan en la superficie del cuerpo del crustáceo. Por otro lado, se encontró que existe relación entre la ocurrencia de los parásitos en los langostinos con la época climática (sequía, lluvia e invernal). De modo que, las condiciones características de cada época climática es determinante para la presencia de los parásitos. Por otro lado, la abundancia de epibiontes sugiere la presencia de altas densidades de bacterias y materia orgánica.

Palabras clave: Abundancia, Endoparásitos, Epibiontes, Langostinos, Pruebas de permutaciones basadas en distancias, PERMANOVA.

1. Introducción

La producción mundial de langostinos comprende principalmente las especies *Macrobrachium rosenbergii* y *M. nipponense*, y en menor escala a *M. lanchesteri*, *M. malcolmsonii* y *M. acanthurus*. Los principales productores son China, Vietnam, India, Tailandia, Bangladesh y Taiwán. *En Latinoamérica, el cultivo de paleamónidos se inició en los años 70's con la especie M. rosenbergii (originaria de Malasia) y se implementó en diversos países del Sur y Centro América incluso en México. Desafortunadamente, la tecnología aplicada no fue la adecuada y la producción esperada no se logró (New & Valenti, 2000).*

En México, la obtención de langostinos proviene esencialmente de la pesca de *M. tenellum* y *M. americanum* en la vertiente del Océano Pacífico, además de *M. acanthurus* y *M. carcinus* en el Golfo de México (Ponce-Palafox *et al.*, 2005). En el estado de Veracruz, se capturan artesanalmente: *M. heterochirus*, *M. acanthurus* y *M. carcinus*, los cuales aportan el 48% de la producción nacional. El langostino *M. acanthurus* cuenta con buena talla comercial (Lorán-Nuñez y Martínez-Isunza, 2002); por lo tanto, un alto potencial para la acuicultura, igualmente tolera los cambios de salinidad, se adapta a bajas concentraciones oxígeno y es menos agresivo que *M. carcinus* (Bowles & Knight, 2000). Sin embargo, poco se sabe sobre los agentes infecciosos que pueden afectar a las poblaciones silvestres de *M. acanthurus*. Es por ello, necesario generar la información científica que permita

establecer la composición de parásitos que pudieran dañar o disminuir la densidad poblacional de langostinos silvestres o de cultivo, lo que permitirá instituir las bases para su manejo o prevención. Por lo tanto, el presente estudio tiene como objeto: a) determinar la ocurrencia de parásitos en poblaciones naturales de langostinos, y b) establecer si las épocas climáticas influyen en la abundancia de parásitos presentes en el hospedero.

2. Metodología

La captura de *M. acanthurus* se realizó en la cuenca baja del río Jamapa, a través de cinco estaciones de muestreo que se ubicaron desde la cabecera municipal de Boca del Río (Longitud 96°06'48"; latitud 19°06' 16") hasta la población de Playa de Vacas, Ver. (Longitud 96°08'50; latitud 19°05' 23"). Durante los meses de octubre de 2007 a julio 2008, se colectaron alrededor de 30 organismos por temporada mediante trampas o clarines. Las temporadas climáticas que se reportan en la región de Veracruz son: invernol (noviembre- febrero), secas (marzo-junio) y lluvias (julio-octubre).

Se realizaron biometrías a los langostinos capturados para determinar el peso (g) y longitud (cm) de cada uno. Posteriormente, se disectaron para extraer las branquias, intestino, tracto digestivo, y hepatopáncreas, los cuales se revisaron bajo microscopio (óptico y estereoscópico), al igual que los apéndices y el exoesqueleto, con la finalidad de analizar los órganos en fresco e identificar a los agentes infecciosos presentes y registrar su ocurrencia. Los parámetros de infección tanto prevalencia y abundancia se obtuvieron de acuerdo a Bush *et al.* (1997), donde se considera a la prevalencia (%) como el promedio de langostinos infectados por cierta especie de parásitos; y a la abundancia, como el número de los parásitos de una especie entre el número total de huéspedes, tanto infectados como no-infectados.

Debido a la naturaleza de datos de la variable abundancia y al no corresponder a una distribución normal, se utilizó un procedimiento propuesto por Anderson *et al.* (2001) de pruebas no paramétricas para comparar a vectores de medias basado en distancias es el siguiente, es decir; se empleó para establecer si existen diferencias entre las abundancias y las temporadas climáticas. Para ello, consideremos a la matriz de datos \mathbf{X} $n \times p$ donde n es el número de elementos y p el número de variables:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

Sea d alguna distancia. A partir de \mathbf{X} se obtiene la matriz de tamaño $n \times n$ $\mathbf{A} = (-d_{ij}^2/2)$ donde d_{ij} es la distancia entre las hileras i y j de \mathbf{X} . Con estas distancias se forma la matriz centrada \mathbf{G} de Gower:

$$\mathbf{G} = \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^T \right) \mathbf{A} \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^T \right)$$

Donde \mathbf{I} es un vector de unos de tamaño $n \times 1$ y \mathbf{I} es la matriz identidad de orden n . Sea \mathbf{D} la matriz $n \times (a - 1)$ de diseño (la matriz que codifica los a grupos que se comparan). Se calcula la matriz de proyección $\mathbf{H} = \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T$ y se calcula el estadístico F para probar la hipótesis de que no hay diferencias entre los a grupos:

$$F = \frac{\text{tr}(\mathbf{HGH})/(a-1)}{\text{tr}((\mathbf{I}-\mathbf{H})\mathbf{G}(\mathbf{I}-\mathbf{H}))/(n-a)}$$

Ahora, se permutan las filas y columnas de \mathbf{G} y se calcula el estadístico de prueba F . Lo anterior se realiza un número grande de veces y con esto se obtiene el valor p . La Tabla 1 muestra algunas distancias en este trabajo.

| Distancia | Expresión |
|-------------|--|
| Euclidiana | $d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$ |
| Orloci | $d_{ij} = \sqrt{\sum_{k=1}^p \left(\frac{x_{ik}}{\sqrt{\sum_{k=1}^p x_{ik}^2}} - \frac{x_{jk}}{\sqrt{\sum_{k=1}^p x_{jk}^2}} \right)^2}$ |
| Ji-cuadrada | $d_{ij} = \sqrt{\sum_{k=1}^p \frac{\sum_{i=1}^n \sum_{k=1}^p x_{ik}}{\sum_{i=1}^n x_{ik}} \left(\frac{x_{ik}}{\sum_{k=1}^p x_{ik}} - \frac{x_{jk}}{\sum_{k=1}^p x_{jk}} \right)^2}$ |
| Canberra | $d_{ij} = \sum_{k=1}^p \frac{ x_{ik} - x_{jk} }{(x_{ik} + x_{jk})}$ |
| Manhattan | $d_{ij} = \sum_{k=1}^p x_{ik} - x_{jk} $ |

Tabla 2.1. Distancias estadísticas utilizadas en este trabajo.

Los análisis no-paramétricos se realizaron en el programa DISTLM V.5.

3. Resultados

Los langostinos capturados tuvieron una longitud total (desde la punta del rostro hasta el telson) entre 4.8 y 12.3 cm y un promedio de 6.66 cm. Los organismos que se capturaron en la época de lluvias fueron significativamente más grandes (longitud y peso) en comparación con los organismos colectados en las otras épocas (lluvias y sequías; $p=0.007$).

| Época climática | Longitud total (cm) LT± D.E. | LT Max | LT Min | Peso (g) P± D.E | Peso Max | Peso Min |
|--------------------|---------------------------------|-----------|-----------|----------------------|-------------|-------------|
| Lluvias (n=30) | 8.31±3.9 ^b | 19 | 4.8 | 4.6±4.2 ^b | 13 | 1.0 |
| Invernal (n=45) | 7.16±2.8 ^a | 19 | 3.8 | 3.5±1.7 ^a | 13 | 0.9 |
| Sequías (n=38) | 7.96±2.07 ^b | 22 | 3.9 | 4.7±4.6 ^b | 15 | 0.8 |

Tabla 3.1. Datos biométricos [longitud total (cm) y peso (g)] de los langostinos

Durante el análisis en fresco, se identificaron cinco protozoarios: *Epistylis* sp., *Lagenophrys* sp., *Acineta* sp., en la superficie del exoesqueleto, apéndices y urópodos, un ciliado apostomado no-identificado en los filamentos branquiales, y dentro las gregarinas en el intestino y hepatopáncreas.

| Distancia | N (p-value) | N (p-value) | N (p-value) | N (p-value) |
|-------------|-------------|-------------|--------------|--------------|
| Euclidiana | 499 (0.002) | 999 (0.001) | 4999(0.0002) | 9999(0.0001) |
| Orloci | 499 (0.002) | 999 (0.001) | 4999(0.0002) | 9999(0.0001) |
| Ji-cuadrada | 499 (0.002) | 999 (0.001) | 4999(0.0002) | 9999(0.0001) |
| Canberra | 499 (0.002) | 999 (0.001) | 4999(0.0002) | 9999(0.0001) |
| Manhattan | 499 (0.002) | 999 (0.001) | 4999(0.0002) | 9999(0.0001) |

Tabla 3.2. P-values obtenidos.

4. Conclusión

Los datos biométricos de los langostinos señalan que la mayoría de los organismos eran juveniles (87%), lo cual señala que el área de estudio es una zona de refugio y criadero para esta especie, así que su conservación y restauración es importante para el mantenimiento de las pesquerías de estos crustáceos en la región.

La composición de parásitos en *M. acanthurus* está integrada por los epibiontes: *Epistylis* sp., *Lagenophrys* sp., *Acineta* sp., y un ciliado apostomado no-identificado en los filamentos branquiales; dentro los endoparásitos se encontró a las gregarinas. Aunque los protozoarios observados no se consideran como agentes infecciosas certificables, en condiciones de estrés, altas densidades de siembra, y/o altos niveles de materia orgánica, los epibiontes registrados pueden causar lesiones en la cutícula de los crustáceos, e interferir en el proceso de muda suprimiendo el crecimiento (Couch, 1983; Kamonporn 1997).

De igual forma, los protozoarios ciliados apostomados en crustáceos provocan la melanización y necrosis, ya que tienen la capacidad para desintegrar la cutícula del hospedero, lo cual lo desprotege contra otros agentes infecciosos oportunistas como bacterias y hongos (Bradbury & Goyal, 1976). Lo anterior, puede disminuir las poblaciones naturales o de cultivo de langostinos, además de dañar la calidad del producto y afectar su comercialización. Por otro lado, las infecciones de epibiontes son un buen indicador de altos niveles de materia orgánica en el agua o cambios en los ecosistemas acuáticos. En el caso de las gregarinas, estos organismos infectan el intestino y hepatopáncreas de los langostinos provocando la destrucción del epitelio intestinal, lo cual afecta la absorción del alimento (Lightner, 1996).

En cuanto análisis estadístico, las pruebas no-paramétricas permitieron establecer que las condiciones específicas que establecen las características particulares de cada temporada climática afectan en la abundancia de parásitos en los langostinos. Este resultado permite determinar de primera instancias, al temporada del año con menos parásitos en *M*

acanthurus, lo cual puede ayudar a establecer épocas de captura para reproductores silvestres con menos abundancia y prevalencia de parásitos.

5. Referencias

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26, 32-46.

Bradbury P.C. & Goyal V. (1976). The fine structure of a parasitic ciliate *Terebrospira* during ingestion of the exoskeleton of a shrimp *Palaemonetes*. *Tissue cell*; 8(43), 573-582.

Bush, A.O., Lafferty K.D., Lotz J.M., & Shostak A.W. (1997). Parasitology meets ecology on its own terms: Margolis *et al.*, revisited. *Parasitology*, 83, 575-583.

Bowles, D.E, Aziz K. & Knight C.L. (2000). *Macrobrachium* (Decapoda: Caridea: Palaemonidae) in the contiguous United States: a review of the species and an assessment of threats to their survival. *Crustacean biology*, 20(1), 158-171.

Couch, J.A. (1983). Diseases caused by Protozoa. In: A.J. Provenzano, (ed.) *The biology of Crustacea: Pathology*. Academy Press London, 79-107

Jayasree L., P. Janakiram & R. Madhavi. (2001). Epibionts and parasites of *Macrobrachium rosenbergii* and *Metapenaeus dobsoni* from Gosthani estuary. *Natural History*, 35(2), 157-167.

Lightner, D.V. (1996). *Handbook of shrimp pathology and diagnostic procedures for diseases of cultured penaeid shrimp*. World Aquaculture Society, Baton Rouge, Louisiana, 1-17

Lorán-Nuñez, RM. y F. R. Martínez-Isunza (2002). Pesquería de langostino. En: Guzmán Amaya P., Quiroga Brahm C., Díaz Luna C., Fuentes Castellanos D., Contreras C. y Silva-López G. (eds.). *La Pesca en Veracruz y sus perspectivas de desarrollo*. SAGARPA. INAPESCA. Universidad Veracruzana, 212-215

New B & W.C. Valenti (2000). Nomenclature y taxonomic. In: B. New and W.C. Valenti (eds.), *Freshwater prawn culture: The farming of *Macrobrachium rosenbergii**. Blackwell Science, Oxford, Inglaterra, 12-17

Ponce-Palafox, J.T., R. González-Salas, O. Romero-Cruz, I. Febrero-Toussaint, J.L. Arredondo-Figueroa, H. Esparza-Leal y G.M. García-Ulloa (2005). Enfermedades del camarón de agua dulce *Macrobrachium tenellum* y *M. rosenbergii* durante el cultivo comercial en estanques rústicos en empresas rurales. *REDVET*, 6(12), 1-12

Impacto del Estrés Materno, Exposición a Plomo, Anemia por Deficiencia de Hierro Pre y Postnatal en el Desarrollo Neuroconductual Infantil

Siliceo-Murrieta, JI¹, Lamadrid-Figueroa, HM², Trejo-Valdivia, B², Mercado-García, A², Wright, RO³, Téllez-Rojo, MM², Wright, R³, Lourdes Schnaas²

¹ Escuela de Salud Pública de México. Instituto Nacional de Salud Pública. Cuernavaca, México.
jose_isaias_simu@yahoo.com.mx

² Centro de Investigación en Salud Poblacional. Instituto Nacional de Salud Pública. Cuernavaca, México

³ Department of Environmental Health. Harvard School of Public Health

Introducción

En las últimas décadas se han realizado diversas investigaciones sobre los factores que impactan en el desarrollo neuroconductual infantil entre los que destacan intoxicación por plomo y deficiencia de hierro durante el embarazo. También se han realizado estudios donde se evalúa el estrés como un factor de riesgo durante el embarazo y que también tienen efectos adversos en la infancia temprana [1].

Es sabido que el plomo es un metal que resulta dañino para los seres humanos, donde los niños son lo más vulnerables a los efectos del mismo ya que se encuentran en proceso de desarrollo.

Los efectos del plomo a la salud durante la infancia inicia desde el primer trimestre de embarazo. Sin embargo, la intoxicación del infante puede continuar los primeros años de vida, por medio de exposiciones ambientales y posiblemente durante la lactancia materna [2]. El desarrollo mental durante la infancia temprana se asocia negativamente con los niveles de plomo en sangre del infante, reflejándose en un decremento del IQ de entre 1 y 4 puntos por un incremento de 10 µg/dL de plomo en sangre [3].

Otro factor que se asocia fuertemente con el efecto del plomo es la presencia de anemia por deficiencia de hierro. Estudios han demostrado que tanto el plomo como la deficiencia de hierro durante el embarazo y la infancia temprana, afectan directa y negativamente a la salud y desarrollo del infante [4].

Sin embargo, el estado emocional de la madre puede ser un factor que modifique el efecto tóxico del plomo al feto, ya que se ha encontrado que altos niveles de estrés durante el embarazo se asocian con bajos puntajes en MDI y PDI así como en el IQ del infante [5].

Estudios previos han considerado al estrés como un confusor del efecto del plomo y de la deficiencia de hierro sobre el neurodesarrollo, sin embargo, la modificación del efecto que pueda ejercer el estrés tiene más sustento biológico que estadístico. Por esta razón, en este estudio se abordó la hipótesis de que el estrés materno tiene un efecto negativo sobre el neurodesarrollo infantil y como modificador de efecto del plomo y deficiencia de hierro

sobre el desarrollo neuroconductual. La relación conceptual propuesta entre los factores ambientales y el desarrollo neuroconductual son representados por la figura 1a.

Materiales y métodos

La población del estudio son las mujeres derechohabientes del Instituto Mexicano del Seguro Social de la Ciudad de México. La muestra se conforma de 133 parejas madre-hijo que aceptaron participar.

Mediciones de los niveles de plomo

En el infante se midió el nivel de plomo en sangre a los 12 y 24 meses de edad. La unidad de medición fue microgramo de plomo por decilitro de sangre ($\mu\text{g/dL}$).

Medición de hemoglobina.

Para evaluar el estado férrico del niño se midió la concentración de hemoglobina y ferritina en sangre, y se tomaron como una variable proxy del nivel de hierro. Las mediciones de hemoglobina y ferritina para el infante fueron tomadas a los 12 y 24 meses de edad. Unidad de medición de hemoglobina y ferritina es gramos por decilitro de sangre (g/dL) y nanogramos por mililitro (ng/mL), respectivamente.

Medición del estrés materno

El estrés se midió mediante una prueba psicométrica que se aplicó a las madres a los 24 meses posterior al parto. El cuestionario estuvo constituido por las escalas de estrés postraumático (PTSD), la escala de percepción de estrés (PSS) y experiencia de vida en pareja (WEB). La confiabilidad de las escalas de estrés empleadas en este estudio fueron evaluadas mediante el alfa de Cronbach y son $\alpha=0.99$ para PTSD, $\alpha=0.62$ para PSS y $\alpha=0.91$ para WEB.

Medición del desarrollo neuroconductual

Para medir el desarrollo neuroconductual se utilizó la escala de desarrollo infantil de Bayley II (BSID II), a los 12 y 24 meses de edad del infante. La escala midió el índice de desarrollo mental (MDI) y el índice de desarrollo motor (PDI).

Análisis estadístico

Las variables MDI y PDI son las variables dependientes primarias y los niveles de plomo, hemoglobina y ferritina infantil y estrés materno representan variables de exposición primaria. Se generaron modelos de regresión lineal múltiple para MDI y PDI evaluando niveles de plomo, hemoglobina y estrés postnatal (3 escalas), en el que se incluyeron las interacciones de segundo orden de las mismas, ajustando por covariables madre e hijo.

Modelo de ecuaciones estructurales (SEM) es una técnica que combina elementos de regresión lineal y análisis de factores. Ésta ofrece la posibilidad de evaluar relaciones e interrelaciones al mismo tiempo, de igual forma el error de medición es considerado en el análisis, donde el verdadero valor de la variable medida con error se define como una variable latente; es decir, se definen la variable latente *plomo postnatal* (valor verdadero) = *plomo en sangre* (valor real) + *error de medición* de plomo en sangre. Los SEM están

formados por dos partes: el modelo de medición y el modelo estructural. El modelo estructural especifica las relaciones causales entre las variables latentes y describe los efectos causales y la cantidad de varianza no explicada. El modelo de medición especifica cómo las variables latentes o constructos hipotéticos son medidos en términos de las variables observadas [6-8].

Podemos escribir el modelo de desarrollo neuroconductual descrito por la figura 1a mediante la siguiente ecuación estructural:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1$$

$$\eta_2 = \beta_{21}\eta_1 + \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \zeta_2$$

donde, η_1 y η_2 son las variables latentes endógenas *estrés materno* y *desarrollo neuroconductual*, respectivamente. ξ_1, ξ_2 son las variables latentes exógenas *concentración de hierro* y *concentración de plomo*, respectivamente. El modelo de medición se define por

$$Y_1 = \lambda_{Y11}\eta_1 + \varepsilon_1$$

$$Y_2 = \lambda_{Y21}\eta_1 + \varepsilon_2 \quad X_1 = \lambda_{X11}\xi_1 + \delta_1$$

$$Y_3 = \lambda_{Y31}\eta_1 + \varepsilon_3 \quad X_2 = \lambda_{X21}\xi_1 + \delta_2$$

$$Y_4 = \lambda_{Y42}\eta_2 + \varepsilon_4 \quad X_3 = \lambda_{X32}\xi_2 + \delta_3$$

$$Y_5 = \lambda_{Y52}\eta_2 + \varepsilon_5$$

donde, las variables observadas son: las escalas de estrés PSS (Y1), PTSD (Y2), WEB (Y3), los índices de desarrollo MDI (Y4) y PDI (Y5), concentración de hemoglobina y ferritina del infante (X1,X2), plomo en sangre del infante (X3),. Los parámetros del modelo fueron estimados por máxima verosimilitud. El ajuste del modelo fue determinado por la prueba ji-cuadrada, y RMSEA. Valores iguales o inferiores a 0.05 para RMSEA e iguales o superiores a 0.90 para la ji-cuadrada son aceptables e indican ajuste razonable entre el modelo y los datos.

Resultados

Modelos de regresión múltiple.

Los resultados de un análisis múltiple para MDI y PDI ajustando por IQ materno, peso y longitud al nacimiento y sexo, se muestran en la Tabla 1. Mediante el ajuste de modelos de regresión, se evaluó la interacción entre estrés materno, concentración de plomo y nivel de hemoglobina del niño. La interacción de tercer orden resultó no significativa por lo que se omitió del análisis. Por el contrario, se encontró asociación significativa ($p < 0.10$) para las interacciones de segundo orden entre la escala de estrés PSS y hemoglobina se asocian positivamente con MDI, en tanto que la interacción plomo hemoglobina se asocia inversamente a MDI. En relación a las otras escalas de estrés no se encontró algún efecto de interacción significativo con plomo o hemoglobina que se asocie a MDI. En lo que se refiere al índice de desarrollo motor (PDI), no se encontró asociación significativa ($p < 0.1$) entre algún efecto de interacción con el índice PDI (Tabla 1).

Tabla 1. Análisis de regresión múltiple del índice de desarrollo mental (MDI) y motor (PDI) de los 24 meses de edad del infante con respecto a las diferentes escalas de estrés materno, nivel de plomo en sangre y de hemoglobina del infante (24 meses) posparto ajustando por IQ materno, peso al nacer del niño, longitud al nacimiento y sexo

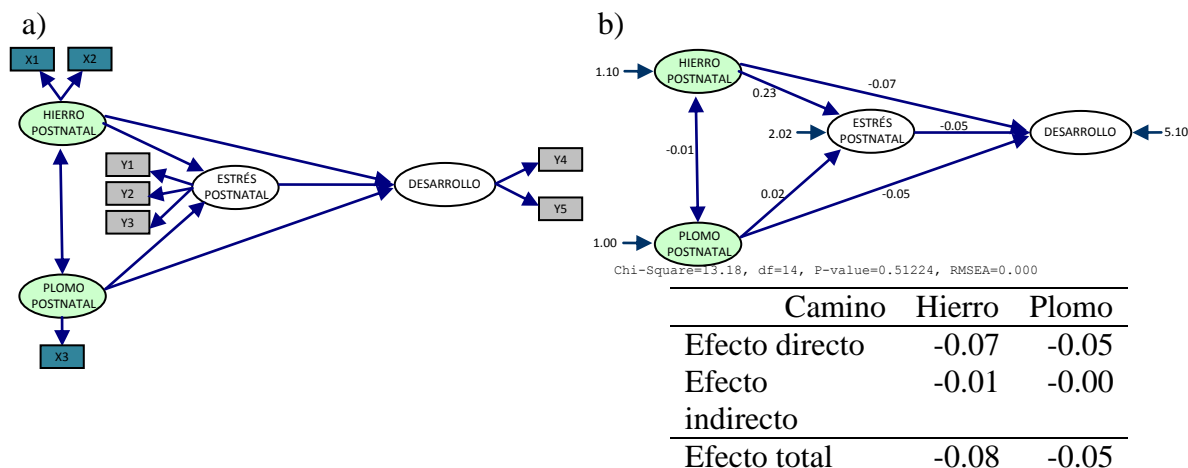
| | PSS | | PTSD | | WEB | |
|--------------------------|-------|---------------|-------|---------------|-------|---------------|
| | Coef. | Valor-p | Coef. | Valor-p | Coef. | Valor-p |
| MDI | | | | | | |
| Escala de estrés | -9.53 | 0.083 | -0.43 | 0.628 | -0.59 | 0.693 |
| Plomo en sangre del niño | 7.73 | 0.057 | 8.75 | 0.047 | 5.08 | 0.282 |
| Hemoglobina en el niño | -4.11 | 0.322 | 2.18 | 0.419 | 2.10 | 0.370 |
| Esc. Estrés*plomo | -0.10 | 0.512 | -0.02 | 0.378 | 0.03 | 0.274 |
| Esc. Estrés*hemoglobina | 0.76 | 0.059 | 0.03 | 0.584 | 0.03 | 0.748 |
| Plomo*hemoglobina | -0.55 | 0.076 | -0.62 | 0.060 | -0.45 | 0.200 |
| Constante | 170.6 | 0.012 | 89.1 | 0.082 | 83.7 | 0.059 |
| R² | | <i>0.1581</i> | | <i>0.1365</i> | | <i>0.1258</i> |
| PDI | | | | | | |
| Escala de estrés | 2.16 | 0.686 | 0.34 | 0.686 | 0.81 | 0.566 |
| Plomo en sangre del niño | 7.18 | 0.070 | 8.59 | 0.040 | 6.04 | 0.180 |
| Hemoglobina en el niño | 3.86 | 0.341 | 3.48 | 0.177 | 3.84 | 0.085 |
| Esc. estrés*plomo | -0.17 | 0.245 | -0.04 | 0.094 | 0.00 | 0.853 |
| Esc. estrés*hemoglobina | -0.12 | 0.749 | -0.02 | 0.786 | -0.06 | 0.549 |
| Plomo*hemoglobina | -0.42 | 0.164 | -0.53 | 0.089 | -0.45 | 0.178 |
| Constante | 10.29 | 0.875 | 22.10 | 0.649 | 13.9 | 0.739 |
| R² | | <i>0.0992</i> | | <i>0.1214</i> | | <i>0.0964</i> |

Modelos de ecuaciones estructurales

El modelo tomó a estrés materno, deficiencia de hierro e intoxicación por plomo en el infante como variables independientes y desarrollo neuroconductual como dependiente. El ajuste de los tres modelos fue aceptable si consideramos que valores iguales o inferiores a 0.05 para RMSEA e iguales o superiores a 0.90 para la ji-cuadrada como adecuados; es decir, el estadístico ji-cuadrado para el modelo presentó un valor de $\chi^2(gl) = 13.18(14)$, $p = 0.51$ y $RMSEA=0.000$

En la figura 1b se representa el diagrama de caminos del modelo ajustado. Se encontró que el efecto de la deficiencia de hierro o el de intoxicación por plomo sobre el desarrollo neuroconductual se modifican cuando se experimenta estrés durante el embarazo. El efecto directo de la deficiencia de hierro del infante sobre el desarrollo es -0.07; esto nos indica que por cada unidad de incremento en el nivel de hierro durante el embarazo representa una disminución de 0.07 unidades en el desarrollo neuroconductual del infante. Sin embargo, dicho efecto se ve modificado cuando se experimenta estrés durante el embarazo, por lo que el efecto de la deficiencia hierro materno disminuye a -0.01 (efecto indirecto). De igual forma ocurre con el efecto de intoxicación por plomo materno, el cual presenta un efecto directo sobre el desarrollo de -0.05 y que podemos interpretar como el incremento en una unidad de intoxicación por plomo implica una disminución de 0.05 unidades en desarrollo neuroconductual del infante. Sin embargo su efecto indirecto a través de estrés se ve modificado a 0.001, es decir, el efecto del plomo pierde fuerza o es mediado por el efecto del estrés.

Figura 1. a) Interrelación entre deficiencia de hierro, intoxicación por plomo y estrés sobre el desarrollo neuroconductual del infante. Variables latentes son hierro, plomo, estrés postnatal y desarrollo neuroconductual. Variables de medición son concentración de hemoglobina, ferritina y plomo en sangre del infante (X1, X2, X3), escala de estrés percibido, estrés postraumático y experiencia de vida en pareja (Y1,Y2,Y3), y finalmente, índice de desarrollo mental y motor (Y4,Y5). b) Modelo ajustado.



Discusión

El hallazgo principal de este análisis es que el efecto del estrés psicosocial durante el embarazo y en los 24 meses posterior al parto, modifica el efecto de la deficiencia de hierro e intoxicación por plomo sobre el desarrollo neuroconductual del infante (MDI y PDI) a los 24 meses de edad ajustando por las covariables IQ materno, peso y longitud al nacimiento y sexo del infante.

Se encontró que el estrés psicosocial materno postnatal (PSS) modifica el efecto del nivel de hemoglobina en el infante sobre el índice de desarrollo mental a los 24 meses de edad. Sin embargo, la interacción entre estrés materno y nivel de plomo en sangre del infante resultó no significativa. Finalmente se encontró una interacción entre los niveles de hemoglobina y plomo sobre el índice de desarrollo mental.

Mediante el ajuste de un modelo estructural con variables latentes encontramos que el efecto directo de la variable latente intoxicación por plomo postnatal se ve mediado cuando éste se evalúa en presencia del estrés postnatal disminuyendo su efecto sobre la variable latente desarrollo neuroconductual. Por otro lado, el efecto de la variable deficiencia de hierro postnatal sobre el desarrollo neuroconductual infantil disminuye su efecto directo cuando el estrés postnatal materno está presente. Tanto deficiencia de hierro como intoxicación por plomo del infante fueron controlados por sus respectivas mediciones maternas durante el embarazo.

Referencias

- 1.- Norbeck, JS., Peterson-Tilden, V. (1983). Life Stress, Social Support, and Emotional Disequilibrium in Complications of Pregnancy: A prospective, Multivariate Study. *Journal of Health and Social Behavior*. **24** 30-46.
- 2.- Hu, H., Téllez-Rojo, MM., Bellinger, D., Smith, D., Ettinger, AS., Lamadrid-Figueroa, H., Schwartz, J., Schnaas, L., Mercado-García, A., and Hernández-Avila, M.(2006). Fetal Lead Exposure at Each Stage of Pregnancy as a Predictor of Infant Mental Development *Environmental Health Perspectives* **114**, **11**.
- 3.- Schnaas, L., Rothenberg, SJ., Flores, MF., Martinez, S., Hernandez, C., Osorio, E., Ruiz Velasco, S. and Perroni, P. (2006). Reduced Intellectual Development in Children with Prenatal Lead Exposure. *Environmental Health Perspectives*. **114**, **5**.
- 4.- Wright, RO., Tsaih, SW., Schwartz, J., Wright, RJ. and Hu, H. (2003). Association Between Iron Deficiency And Blood Lead Level In A Longitudinal Analysis Of Children Followed In An Urban Primary Care Clinic. *J Pediatrics*. **142:9-14**.
J. Nutr. 134:363-371, February 2004.
- 5.- Laplante, DP., Barr, RG., Brunet, A., Du Fort, GG., Meaney, ML., Jean-Franc, Saucier, O., Zelazo, PR. and King, S. (2004). Stress During Pregnancy Affects General Intellectual and Language Functioning in Human Toddlers. *Pediatric Research*. International Pediatric Research Foundation, Inc. **56**, **3**.
- 6.- Budtz-Jorgensen, E., Keiding, N., Grandjean, P. and Weihe, P. (2002). Estimation of health effects of prenatal methylmercury exposure using structural equation models. *Environmental Health: A Global Access Science Source*. **1**.
- 7.-Sanchez, B., Budtz-Jorgensen, E., Ryan, LM. and Hu, H. (2005). Structural Equation Models: A Review With Applications to Environmental Epidemiology. *Journal of the American Statistical Association*. 1443-1455.
- 8.-Schermelleh-Engel, K. and Moosbrugger, H. (2003). Evaluating the Fit of Structural Equation Models: Test of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*. **8**. 23-74.

Estimación de Estadísticos de Primer y Segundo Orden Basados en el Análisis de Imágenes Digitales de la Naranja Valencia

Barradas-Domínguez P.¹, Jiménez-Fernández M.², Atanasio-Gómez R.³

¹Instituto Tecnológico Superior de Xalapa. pbarrada@hotmail.com

²Instituto de Ciencias Básicas. Universidad Veracruzana. marijife@hotmail.com

³Instituto Tecnológico Superior de Xalapa. reyes010587@hotmail.com

Resumen

El presente trabajo describe la implementación de un sistema de procesamiento digital de imágenes para analizar los parámetros de color y características texturales de cítricos implementando el cálculo de 6 estadísticos de primer y 6 de segundo orden derivados del histograma en escala de grises y de la matriz de coocurrencia. Se utilizaron fotografías de un lote de 26 naranjas variedad valencia para el experimento obteniendo resultados exitosos al poder correlacionar los parámetros calculados con las propiedades físicas de la naranja como lo son el color y textura y así poder realizar una clasificación de los cítricos.

1. Introducción

Un atributo de una imagen es una característica primitiva con la cual se puede distinguir. Algunas características son naturales como las del sentido tal como las características definidas por la apariencia visual de la imagen, mientras otras, son características artificiales resultado de la manipulación específica de una imagen (William, 2001). El pixel es el componente básico de las imágenes. Dos tipos de información se encuentran en cada pixel; el valor de la intensidad luminosa y las coordenadas que están asignadas a la imagen. La textura tiene que ver con ambas; la intensidad luminosa y las coordenadas que están asignadas a la imagen (Zheng et al, 2006). Existe una variedad de medidas texturales que van de simples estadísticas como la desviación estándar a fractales, campos aleatorios de Markov, funciones de Gabor, longitud de onda, estadísticos derivados de la matriz de coocurrencia, entre otros (Berberoglu et al, 2007). La textura de una imagen representa una variación espacial en el espectro de la brillantez de la imagen digital. Las medidas texturales son usualmente extraídas de los principales componentes de radiación en un conjunto de bandas de ondas (Berberoglu et al, 2007) y pueden ser aplicadas en la industria de alimentos para la evaluación de la calidad y la inspección (Zheng et al, 2006) ya que la superficie de la textura provee contribuciones importantes para analizar un producto alimenticio (Qiao et al, 2007) y puede en cierta medida reflejar la estructura celular de los productos alimenticios y utilizarse como un indicador de la calidad de los alimentos (Zheng et al, 2006).

En análisis de imágenes hay un número de técnicas para cálculo de propiedades texturales de una imagen. Estas son usualmente categorizadas en métodos estáticos, espectrales y estructurales (Fernández et al, 2005). Entre todos los métodos estáticos, el más popular es uno basado en la estimación de estadísticos de primer y segundo orden. En la industria de los alimentos los métodos estáticos para calcular las características texturales son los más usados para determinar los niveles de calidad o para la clasificación. El enfoque estático trata de caracterizar la textura de una región de una imagen usando medidas estáticas. Un histograma de primer orden representa la distribución de las intensidades de los niveles de

grises. Los parámetros de los estadísticos de segundo orden tal como la matriz de co-ocurrencia en niveles de grises pueden extraer información textural de la distribución espacial y la interrelación de los niveles de grises (Haruka et al, 2007).

El objetivo de este estudio fue implementar métodos para el análisis textural de imágenes digitales de naranjas variedad valencia basados en el cálculo de estadísticos de primer y segundo orden para establecer una correlación entre los valores de parámetros estadísticos con las características sensoriales de color y textura para la clasificación de la naranja variedad valencia.

2. Materiales y Métodos

Materia Prima

Las naranjas utilizadas como muestra se adquirieron en un mercado local de la región de Xalapa, Ver. Se seleccionaron 26 naranjas (variedad valencia) con diferentes tonalidades de color las cuales fueron de la más verde a la más anaranjada. Además se tomó en cuenta que las naranjas tuvieran diferentes niveles de rugosidad en la cascara y diferentes tamaños (figura 1).

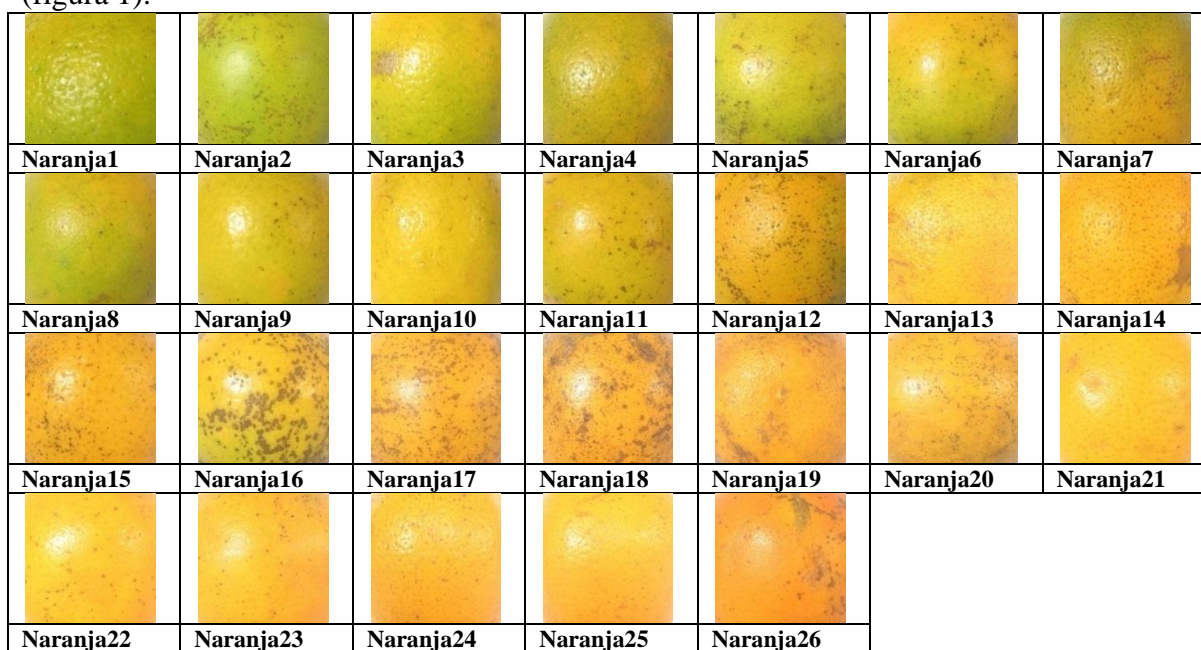


Figura 1. Lote de 26 naranjas (variedad valencia) el cual fue utilizado como muestra para las pruebas. Se puede apreciar las diferentes tonalidades de cada una, las cuales van de la más verde a la más anaranjada.

Cámara Digital y Adquisición de la Imagen

Una cámara digital (CCD), modelo *PowerShot A640 (Canon, USA 2006)* de 10.0 mega píxeles fue colocada verticalmente sobre el fondo donde se colocó la muestra a una distancia de 22 cm. Como condiciones estándar de captura, las imágenes fueron tomadas sobre un fondo blanco con las siguientes configuraciones de la cámara; modo manual, sin zoom, sin flash, resolución de la imagen (3648 x 2736) píxeles, y almacenadas en formato JPEG. La cámara fue conectada por el puerto USB a una laptop *Gateway* modelo 6846 de arquitectura 64 bit con 4 gigabytes de memoria RAM y procesador *Intel Core2 Duo* para poder adquirir y visualizar directamente las imágenes.

Software

Todo los algoritmos de pre-procesamiento, segmentación, extracción de características internas, cálculo de los estadísticos de primer y segundo orden, fueron implementados en el lenguaje de programación C con el compilador *gcc* 4.1.2 y librerías gráficas OpenGL-Mesa 7.3. El sistema operativo fue Slackware-Linux versión 12.1.

Pre-procesamiento de la Imagen

Las imágenes originales tomadas por la cámara CCD a color fueron transformadas a imágenes monocromáticas (a escala de grises). La conversión a escala de grises de la imagen es esencial para realizar la implementación de los algoritmos de análisis textural, así que primero es necesario realizar este procedimiento. Para esto la imagen se recorrió completamente, calculando en cada píxel la intensidad de gris que le corresponde acorde a la siguiente ecuación de iluminación (Qing et al, 2007), en donde I_r , I_g y I_b corresponden a los valores de las intensidades de los componentes rojo, verde y azul, respectivamente.

$$I = 0.3I_r + 0.59I_g + 0.11I_b$$

Extracción de Características Texturales

Los estadísticos de primer orden; media, desviación estándar, skewness, kurtosis, energía y entropía fueron calculados a partir del histograma de la imagen en escala de grises. La forma del histograma de una imagen provee muchas pistas como para determinar la forma esta. En la estimación de estadísticos de segundo orden cada característica textural es computada a un conjunto de elementos de una matriz de coocurrencia (COM por sus siglas en ingles *co-occurrence matrix*) que contiene la relación entre la intensidad de un pixel y la de sus pixeles vecinos (Fernández et al., 2005). Una matriz de coocurrencia (COM) es una matriz cuadrada cuyos elementos corresponden a la frecuencia relativa de ocurrencia $P(i, j, d, \theta)$ de los valores de dos pixeles (uno con intensidad i y el otro con intensidad j), separados por una cierta distancia d en una dirección dada θ . Por lo tanto una COM es una matriz cuadrada que tiene en tamaño del valor del pixel cuyo valor es el más alto en toda la imagen. De este modo el histograma 2D puede representar algunas características importantes de una textura de la imagen (Gao y Tan, 1996). Haralick et al., 1973 citados por Fernández et al., 2005 propusieron 14 medidas de características texturales derivadas de la COM, cada una representando una propiedad específica de la imagen como aspereza, homogeneidad del contraste, complejidad de la textura, entre otras. Las características no son independientes una de otra y algunas indican propiedades similares de la imagen (Fernández et al., 2005). Únicamente 6 de las 14 características texturales que están extensamente usadas en la literatura fueron seleccionadas las cuales fueron; covarianza, inercia, valor absoluto, diferencia inversa, energía y entropía.

3. Resultados y Discusiones

Se realizaron pruebas con imágenes sintéticas de diferentes tipos de texturas generadas de manera aleatoria controlando solo los diferentes niveles de intensidad de color obteniendo resultados satisfactorios. A medida que la textura de las imágenes es más homogénea los valores de algunos parámetros incrementan mientras que otros se comportan de manera contraria al mostrar un decremento. Así pues con respecto a los estadísticos de primer orden la entropía mostrando valores de 0 a 8, la media con posibles valores de 0 a 255, y la desviación estándar, decrecen su valor cuando la textura tiende a ser más uniforme, la

energía mostrando valores de 0 a 1 ,y la kurtosis incrementan su valor a medida que la textura es más uniforme, los valores que toman la media, la desviación estándar y la kurtosis dependen del número de pixeles analizados así como de los valores máximos y mínimos de la intensidad de gris encontrados durante todo el análisis de la imagen completa, el parámetros skewness fue el único que mostro valores aleatorios a lo largo de todas las muestras. De manera similar que en los estadísticos de primer orden, para los estadísticos de segundo orden los valores en algunos parámetros incrementan y otros decrecen conforme la textura analizada es más homogénea, la entropía tomando valores del doble de la entropía de primer orden, la media teniendo los mismos valores que la media de primer orden, la inercia y el valor absoluto decrecen cuando la textura es más uniforme, en contraposición la energía que tomo valores de la energía de primer orden elevados al cuadrado, y la diferencia inversa la cual también toma valores de 0 a 1 incrementan su valor para texturas mas homogéneas.

Se realizo el mismo cálculo para las imágenes de cada una de las naranjas utilizadas como muestras en el experimento, mostrando resultados exitosos ya que para cada unos de los parámetros cuyo valor tienden a incrementarse cuando se analiza una textura homogénea, en el análisis de las imágenes correspondientes a las naranjas mas rugosas de acuerdo al punto de vista del observador humano estos parámetros efectivamente mostraron valores altos, por otro lado los parámetros que mostraron valores bajos y con un patrón de comportamiento de decremento fueron aquellos que en la prueba con las imágenes sintéticas decrecían a medida que la textura era más uniforme. La figura 1 muestra los valores obtenidos para el estadístico de primer orden entropía correspondiente a cada una de las imágenes del lote de naranjas, la grafica indica que las naranjas cuya textura es más homogénea corresponden a las naranjas 21 a la 26 siendo las naranjas 21 y 22 las que muestran una textura más uniforme de acuerdo a que son las que muestran los valores más altos para este estadístico, los valores más bajos de esta grafica son aquellos que corresponden a las naranjas con una textura mas rugosa que el resto de la muestra, las cuales son la naranja 5, 12 y 16. Los resultados obtenidos para la desviación estándar se ilustran en la figura 2, para el caso de este estadístico el comportamiento es inverso al de la energía ya que este parámetro indica una textura uniforme con valores bajos mientras que las texturas rugosas están dadas por valores altos en la desviación estándar, de acuerdo a la grafica de la figura 2 se puede determinar que las naranjas cuya textura es más uniforme son las naranjas de la 21 a la 26 las cuales muestran los valores más bajos de toda la muestra (caso contrario de la energía) así mismo podríamos determinar que las naranjas 12 y 16 corresponden a las naranjas que tienen la textura mas rugosa.

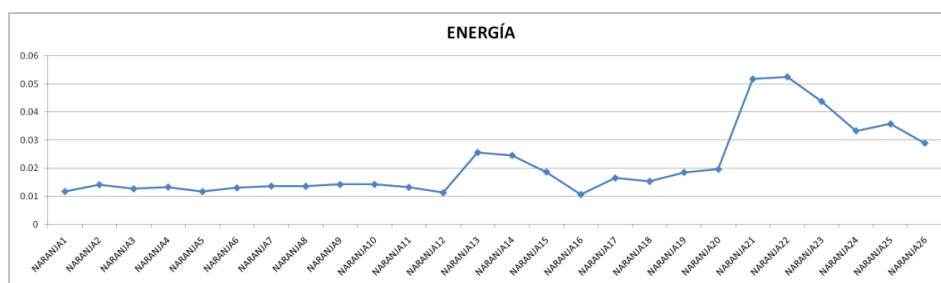


Figura 2. Estadístico de primer orden energía calculado para cada una de las imágenes del lote de naranjas utilizado como muestra.



Figura 2. Estadístico de primer orden desviación estándar calculado para cada una de las imágenes del lote de naranjas utilizado como muestra.

Los resultados mostrados en las graficas se pueden verificar inspeccionando visualmente las muestras de la figura 1.

4. Conclusiones

La aplicación de métodos estadísticos en el procesamiento digital de imágenes es una herramienta muy poderosa con la cual es posible discriminar texturas para llevar a cabo clasificación de cítricos y de esta manera la utilización de técnicas no destructivas para el control de calidad en diversos procesos.

REFERENCIAS

- Berberoglu, S., Curran, P.J., Lloyd, C.D., Atkinson P.M. (2007) Texture Classification of Mediterranean Land Cover. *International Journal of Applied Earth Observation and Geoinformation* **9**, 322–334.
- Fernández, L., Castellero, C., Aguilera, J.M. (2005) An application of image analysis to dehydration of apples discs. *Journal of Food Engineering* **67**, 185–193.
- Haruka, D., Teruaki, A., Karou, K. (2007) Characterization of spatiotemporal stress distribution during food fracture by image texture analysis methods. *Journal of Food Engineering* **81**, 429–436.
- Qiao, J., Wang, N., Ngadi, M.O., Kazemi, S. (2007) Predicting mechanical properties of fried chicken nuggets using image processing and neuronal network techniques. *Journal of Food Engineering* **79**, 1065–1070.
- Qing, Z., Ji, B., Zude, M. (2007) Predicting soluble solid content and firmness in apple fruit by means of laser light backscattering image analysis. *Journal of Food Engineering* **82**, 58–67.
- William, K. (2001). *Digital Image Processing*. California: Wiley-Interscience.
- Zheng, C., Sun, D.W., Zheng, L. (2006) Recent applications of image texture for evaluation of food qualities – a review. *Trends in Food Science & Technology* **17**, 113–128

Funciones de Varianza, Covarianza y Correlación Bicuadrática para Distribuciones Normales

Carlos Eduardo Alonso Malaver¹, Jorge Martínez Collantes²

¹Universidad Nacional de Colombia sede Bogotá. Departamento de Estadística.

cealonsom@unal.edu.co

²Universidad Nacional de Colombia sede Bogotá. Departamento de Estadística.

jmartinezc@unal.edu.co

Resumen

En este trabajo se analiza el comportamiento del coeficiente de correlación bicuadrático ρ , asumiendo que se tienen dos variables aleatorias con distribución normal bivariada. Coeficiente desarrollado a partir de las ideas de BEATON & Tukey (1974) y LAX (1975). Los principales desarrollos se hacen con base en los trabajos de COHEN (1955), KHATRI & Jaiswal (1963) y ROSENBAUM (1961) sobre distribuciones normales truncadas. E indican que el coeficiente de correlación ρ toma valores muy cercanos al valor de ρ .

De acuerdo a los resultados $\rho > \rho$ cuando $\rho < 0$, $\rho = 0$ cuando $\rho = 0$ y $\rho < \rho$ cuando $\rho > 0$. De lo anterior el comportamiento que se debe esperar del estimador bicuadrático de correlación es que sobreestima el valor de ρ cuando la correlación es negativa, y subestima el valor de ρ cuando se tiene asociación positiva, es decir no es insesgado.

El reto de largo plazo, y que no culmina con este trabajo, es plantear estimadores robustos insesgados o asintóticamente insesgados del coeficiente de correlación ρ y de la función de autocorrelación ρ_h , a partir de los estimadores planteados en los trabajos de TUNJANO (2006), RICAURTE & Martínez (2007) y VALCARCEL (2007).

Palabras clave: Estadística Robusta, Estimadores M, Coeficiente de Correlación, Proceso Estocástico Estacionario, Variables Acotadas.

1. Estimador de Correlación Bicuadrático

El estimador clásico de la función de autocorrelación (FAC) parámetro¹, es una herramienta central en el análisis de series de tiempo, porque a partir de ella se identifica el modelo, se construyen los estimadores de los parámetros, se analizan los residuales, entre otros. Pero en contraste con esa gran importancia, este estimador es altamente sensible ante la presencia de datos atípicos ver WEI (2006) y VALCARCEL (2007).

A partir de lo anterior se planteó el estimador bicuadrático de correlación como un estimador robusto de ρ_h . Planteamiento que se hace a partir de las ideas de LAX (1975). El estimador propuesto, ver TUNJANO (2006) y RICAURTE & Martínez (2007), es de la forma,

$$\hat{\rho}_{xy} = \frac{\delta_{xy}}{\delta_{xx}\delta_{yy}}, \quad (1)$$

¹E-mail: cealonsom@unal.edu.co

²E-mail: jmartinezc@unal.edu.co

¹ $\hat{\rho}_h = r_h = \frac{\hat{\gamma}_h}{\gamma_0}$ con $\hat{\gamma}_h = \frac{\sum_{i=1}^{n-|h|} (x_i - \bar{x})(x_{i+|h|} - \bar{x})}{n}$. Se divide por n , para asegurar que las matrices de autocovarianza y autocorrelación sean definidas positivas, BROCKWEL & Davis (1991), página 29.

con δ_{xy} esta dado por,

$$\delta_{xy} = n(kMAD_x)(kMAD_y) \frac{\sum_{i=1}^n \psi(z_{x_i})\psi(z_{y_i})}{\left(\sum_{i=1}^n \psi'(z_{x_i})\right) \left(\sum_{i=1}^n \psi'(z_{y_i})\right)}, \quad (2)$$

Donde k es una constante positiva de estandarización², $z_{x_i} = \frac{x_i - \bar{x}}{kMAD_x}$, $z_{y_i} = \frac{y_i - \bar{y}}{kMAD_y}$, $\bar{x} = \text{mediana}\{x_1, x_2, \dots, x_n\}$, $MAD_y = \text{Mediana}|x_j - \bar{x}|$, y $\psi(z) = 1_{\{|z|<1\}} z(1-z^2)^2$ ($1_{\{\cdot\}}$ variable indicadora).

2. Resultados

2.1. Generalización de la Covarianza

Dadas dos variables aleatorias X y Y , con función distribución conjunta y segundos momentos univariados finitos ($E(X^2) < \infty$ y $E(Y^2) < \infty$), la covarianza se define como:

$$\gamma_{XY} = E[(X - EX)(Y - EY)] \quad (3)$$

ecuación que se puede generalizar, a partir de las ideas en LAX (1975), como

$$\gamma_{XY} = \frac{E[\psi(Z_x)\psi(Z_y)]}{E[\psi'(Z_x)]E[\psi'(Z_y)]}$$

Donde $\psi(x) = x$, $\psi'(x) = \frac{\partial \psi(x)}{\partial x}$, $Z_x = X - EX$ y $Z_y = Y - EY$. Resultado a partir del cual se plantea una generalización de la covarianza.

Definición 1 (ψ -Covarianza). Dadas dos variables aleatorias X y Y , con función distribución conjunta y segundos ψ -momentos finitos ($E[\psi^2(X)] < \infty$ y $E[\psi^2(Y)] < \infty$)³, la ψ -covarianza entre X e Y se define como:

$$\gamma_{\psi XY} = \frac{E[\psi(Z_x)\psi(Z_y)]}{E[\psi'(Z_x)]E[\psi'(Z_y)]} \quad (4)$$

Definición 2 (*Coefficiente de ψ -Correlación*). Dadas dos variables aleatorias X y Y , con $\gamma_{\psi XX} < \infty$ y $\gamma_{\psi YY} < \infty$, el coeficiente de ψ -correlación entre X e Y se define como:

$$\rho_{\psi XY} = \frac{\gamma_{\psi XY}}{(\gamma_{\psi XX} \gamma_{\psi YY})^{\frac{1}{2}}} \quad (5)$$

Se puede mostrar que el coeficiente $\rho_{\psi XY}$ planteado en la definición 2, es el funcional asociado al estimador $\hat{\rho}_{\psi XY}$, ecuación (1). En este trabajo se estudia el comportamiento del funcional $\rho_{\psi XY}$ buscando las posibles causas del no insesgamiento de $\hat{\rho}_{\psi XY}$, ver TUNJANO (2006), desarrollo que se hace asumiendo se tienen dos variables aleatorias X y Y con distribución normal bivariada.

²La propuesta de Lax es $k = 9$, valor planteado desde las propiedades de una distribución $N(\mu, \sigma^2)$, donde $MAD \approx \frac{2}{3}\sigma$, así $kMAD = 9MAD \approx 6\sigma$, HOAGLIN et al. (1983), página 417.

³Esto se plantea inicialmente para polinómicos.

2.2. Varianza Bicuadrática

Si se tienen las variables aleatorias X y Y con distribución normal bivariada⁴ $N(0, 0, 1, 1, \rho)$, las distribuciones univariadas para X y Y , son del tipo $N(0, 1)$, de donde se tiene $Z_x = \frac{X - M_{AD}X}{kM_{AD}X} = \frac{X}{1}$ y $Z_y = \frac{Y - M_{AD}Y}{kM_{AD}Y} = \frac{Y}{1}$, con $l = k \times 0.67448975$, con k constante de estandarización planteada por Lax. Si se define la variable M como $M = 1_{\{|X| < l\}} X$ ($1_{\{\cdot\}}$ es una variable indicadora), la varianza bicuadrática se puede escribir como:

$$\gamma_{\psi_{XX}} = \frac{E_x - 4\frac{E_x^2}{l^2} + 6\frac{E_x^3}{l^3} - 4\frac{E_x^4}{l^4} + \frac{E_x^5}{l^5}}{(1 - 6\frac{E_x^2}{l^2} + 5\frac{E_x^3}{l^3})^2}, \quad (6)$$

donde $E_x = E(M^x)$. La función de densidad de M esta dada por $f_M(m) = \frac{\phi(m)}{c_1} 1_{\{|m| < l\}}$. La función $f_M(\cdot)$ es la función de densidad resultante de truncar a dos colas, la función de densidad asociada a una distribución normal estándar $\phi(\cdot)$, y $c_1 = P(-l < X < l)$.

Si se define $m_M^{(r)}(t) = \frac{\partial^r E(e^{tM})}{\partial t^r}$, el comportamiento de la derivada de orden $k \geq 2$ de la función generatriz de M se puede generalizar, como,

$$m_M^{(r)}(t) = (r-1)m_M^{(r-2)}(t) + tm_M^{(r-1)}(t) + (-l)^{r-1} \frac{e^{-\frac{1}{2}t^2} [e^{-lt} + (-1)^k e^{lt}]}{c_1(2\pi)^{\frac{1}{2}}} \quad \text{para } r \geq 2 \quad (7)$$

§Cálculo de Varianza Bicuadrática y Varianza de Lax

A partir de la ecuación (4) para el caso $X = Y$ y la ecuación (7), es posible calcular los funcionales asociados a la varianza bicuadrática. Los resultados se muestran en la Tabla 1.

TABLA 1: Valores de la Varianza Bicuadrática para una Distribución Normal.

| | Valor de k | | |
|----------------------|----------------|----------------|----------------|
| | 3 | 6 | 9 |
| $\gamma_{\psi_{XX}}$ | 0.491377330220 | 0.066819506434 | 0.027637604055 |

2.3. Coeficiente de Correlación Bicuadrático

Si se define el vector aleatorio $\mathbf{M} = (M_1, M_2)^T$ como $\mathbf{M} = 1_{\{|X| < l, |Y| < l\}} \mathbf{X}$, con $\mathbf{X} = (X, Y)$ vector aleatorio con distribución $N(0, 0, 1, 1, \rho)$, la función bicuadrática en terminos de las componentes del vector $\mathbf{M} = (M_1, M_2)^T$, se puede escribir como;

$$\gamma_{\psi_{XY}} = \frac{\frac{E_{1,1}}{l^2} - 4\frac{E_{3,1}}{l^4} + 2\frac{E_{5,1}}{l^6} + 4\frac{E_{3,3}}{l^6} - 4\frac{E_{5,3}}{l^8} + \frac{E_{5,5}}{l^{10}}}{\left[1 - 6\frac{E_{M_1,2}}{l^2} + 5\frac{E_{M_1,4}}{l^4}\right]^2} \quad (8)$$

donde $E_{r,h} = E(M_1^r M_2^h)$, $E_{M_1,r} = E(M_1^r)$ y $E_{M_2,r} = E(M_2^r)$, para r y h enteros⁵. Análogo a lo realizado para el caso de la varianza bicuadrática, el vector aleatorio \mathbf{M} , es resultado de truncar un vector con

⁴La notación se emplea en el siguiente sentido, dadas dos variables aleatorias X y Y , con distribución normal bivariada con vector de medias $\mu = (\mu_1, \mu_2)^T$ y matriz de varianzas y covarianzas dada por;

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

Se dice que X y Y , tienen distribución normal bivariada $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

⁵ $E(M_1^r) = E(M_2^r)$, dado que X e Y tienen la misma distribución

distribución normal, su función de densidad conjunta esta dada por $f_{\mathbf{M}}(m_1, m_2) = \frac{\phi(m_1, m_2)}{c_2} 1_{\{|m_1| < l, |m_2| < l\}}$, con $\phi(\cdot, \cdot)$ función de densidad conjunta $N(0, 0, 1, 1, \rho)$, y $c_2 = P(-l < X < l, -l < Y < l)$.

§ **Covarianza bicuadrática para variables aleatorias independientes**

Si se supone $\rho = 0$, de la definición del vector $\mathbf{M} = (M_1, M_2)$, se sigue que las variables aleatorias M_1 y M_2 , son independientes. Resultado a partir del cual la ecuación (8) se transforma en

$$\gamma_{\psi_{XY}/\rho=0} = \frac{\frac{E^2(M_1)}{l^2} - 4\frac{E(M_1^2)E(M_1)}{l^4} + 2\frac{E(M_1^2)E(M_1)}{l^4} + 4\frac{E^2(M_1^2)}{l^4} - 4\frac{E(M_1^2)E(M_1^2)}{l^6} + \frac{E^2(M_1^2)}{l^6}}{\left[1 - 6\frac{E_{M_1,2}}{l^2} + 5\frac{E_{M_1,4}}{l^4}\right]^2} \quad (9)$$

El denominador de esta expresión es distinto de cero, y en el numerador sólo se tienen momentos de orden impar de la variable M_1 , dado que la función de densidad de esta variable es simétrica alrededor de cero, se tiene $\gamma_{\psi_{XY}/\rho=0} = 0$.

§ **Covarianza bicuadrática - $\rho \neq 0$**

El trabajo con distribuciones normales truncadas no es nuevo, Pearson inicialmente trabajo en los 1930's sobre estas distribuciones, con el propósito de generar algunas tablas⁶; posteriormente trabajan sobre normales truncadas COHEN (1955), SINGH (1960), ROSENBAUM (1961), TALLIS (1961), FINNEY (1962) y KHATRI & Jaiswal (1963). Si se nota $\mathbf{t} = (t_1, t_2)^T$, la función generatriz de momentos del vector aleatorio \mathbf{M} , esta dada por,

$$G_{\mathbf{M}}(\mathbf{t}) = E\left(e^{\mathbf{t}^T \mathbf{M}}\right) = \frac{e^{\frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}}}{c_2} \int_{-l}^l \int_{-l}^l \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)} e^{-\frac{1}{2}(\mathbf{m} - \Sigma \mathbf{t})^T \Sigma^{-1}(\mathbf{m} - \Sigma \mathbf{t})} dm_1 dm_2 \quad (10)$$

Los momentos conjuntos de orden r y h de \mathbf{M} (r y h enteros nonegativos), alrededor del origen, están dados por,

$$E(M_1^r M_2^h) = \frac{\partial^{r+h}}{\partial t_1^r \partial t_2^h} G_{\mathbf{M}}(\mathbf{t}) \Big|_{t_1=0, t_2=0} \quad (11)$$

§ **Valores de $\varrho_{\psi_{XY}}$**

A partir de la ecuación (11), se llega a los momentos conjuntos del vector \mathbf{M} y con estos al valor de $\gamma_{\psi_{XY}}$, ecuación (4). Valores a partir de los cuales es posible calcular el coeficiente de correlación bicuadrático, ecuación (7). Los valores de $\gamma_{\psi_{XX}}$, ya habían sido obtenidos, Tabla 1.

Los resultados muestran que $\varrho_{\psi_{XY}}(\rho) \rightarrow \rho$ cuando k crece. Para $k = 9$, los valores de $\varrho_{\psi_{XY}}(\rho) \approx \rho$, ver Tabla 2. El comportamiento de este funcional sugiere que el estimador bicuadrático $\hat{\varrho}$, subestima el valor de ρ , cuando $\rho > 0$, y sobreestima su valor cuando $\rho < 0$.

Referencias

BEATON, A. & Tukey, J. (1974), 'The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data', *Technometrics* **16**(2), 147-185.
 BROCKWEL, P. & Davis, R. (1991), *Time Series: Theory and Methods*, Springer Verlag, New York.
 COHEN, C. (1955), 'Restriction and selection in samples from bivariate normal distributions', *Technometrics* **50**(271), 884-893.

⁶Tomado de ROSENBAUM (1961), página 405.

- FINNEY, D. (1962), 'Cumulants of truncated multi-normal distributions', *Journal of the Royal Statistical Society* **24**(2), 535-536.
- HOAGLIN, D., Mosteller, F. & Tukey, J. (1983), *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, New York.
- KHATRI, C. & Jaiswal, M. (1963), 'Estimation of parameters of a truncated bivariate normal distribution', *JASA* **58**(302), 519-526.
- LAX, D. (1975), An interim report of a monte carlo study of robust estimators of withers, Series 2 - 93, Princeton University, Department of Statistics.
- LAX, D. (1985), 'Robust estimators of scale: Finite sample performance in log-tailed symmetric distributions', *JASA* **80**(391), 736-741.
- RICAURTE, N. & Martínez, J. (2007), Estimacion de la funcion de probabilidad del coeficiente de correlacion basado en la funcion bicuadratica. Departamento de Estadística - Universidad Nacional.
- ROSENBAUM, S. (1961), 'Moments of a truncated bivariate normal distribution', *Journal of the Royal Statistical Society* **23**(2), 405-408. Serie B.
- SINGH, N. (1960), 'Estimation of parameters of a multivariate normal population from truncated and censored samples', *Journal of the Royal Statistical Society* **22**(2), 307-311. Serie B.
- TALLIS, G. (1961), 'The moment generating function of the truncated multinormal distribution', *Journal of the Royal Statistical Society* **23**(1), 223-229. Serie B.
- TUNJANO, W. (2006), Propuesta de un estimador de covariabilidad basado en la función bicuadrática. Departamento de Estadística - Universidad Nacional.
- VALCARCEL, H. (2007), Propuesta de una función de autocorrelación con base en la función bicuadrática. Departamento de Estadística - Universidad Nacional.
- WEI, W. (2006), *Time Series Analysis, Univariate and Multivariate Methods*, ii edn, Pearson Addison Wesley.

TABLA 2: Valores de ρ de Acuerdo a los Valores de ρ y k

| Valor de ρ | Valor de k | | |
|-----------------|--------------|----------|----------|
| | 3 | 6 | 9 |
| 0.000 | 0.000000 | 0.000000 | 0.000000 |
| 0.100 | 0.049966 | 0.091461 | 0.098207 |
| 0.200 | 0.101565 | 0.183422 | 0.196523 |
| 0.300 | 0.156766 | 0.276384 | 0.295054 |
| 0.400 | 0.218223 | 0.370858 | 0.393911 |
| 0.500 | 0.289691 | 0.467366 | 0.493201 |
| 0.600 | 0.376515 | 0.566446 | 0.593034 |
| 0.700 | 0.486181 | 0.668655 | 0.693520 |
| 0.800 | 0.628749 | 0.774574 | 0.794769 |
| 0.900 | 0.815515 | 0.884808 | 0.896891 |
| 0.999 | 1.010647 | 0.998821 | 0.998964 |

Inferencia Sobre Datos No Detectados de Contaminantes

Dr. Fidel Ulín Montejo¹, Dr. Humberto Vaquera Huerta²

¹Universidad Juárez Autónoma de Tabasco. fidel.ulín@basicas.ujat.mx

²Colegio de Postgraduados. hvaquera@colpos.mx

1. Introducción

Con la finalidad de conocer el riesgo de los contaminantes con niveles extraordinariamente bajos, como los metales pesados, se buscan medirlos y cuantificarlos; sin embargo, a veces la señal producida por el contaminante es tan pequeña que los instrumentos no pueden cuantificarla, reportándose como datos no-detectados (Millard y Neerchal, 2001; Helsel, 2005). Tales observaciones son datos censurados por la izquierda. El punto de referencia de la censura es el límite de detección. En (Gilbert, 1987; Millard y Deverel, 1988) se muestran procedimientos para comparar poblaciones a través de la mediana; donde los datos no-detectados son omitidos o sustituidos arbitrariamente; en ambos casos los alcances y resultados son limitados. En algunos casos la normalidad se asume con la log-transformación de los datos, sin embargo, como se discute en (El-Shaarawi y Viveros, 1997), los resultados necesitan ser reportados sobre las mediciones obtenidas y su media, mas que en la escala log-transformada y su mediana, como lo ha establecido la EPA (1992). En este sentido, los datos ambientales presentan características que conducen a problemas interesantes, donde la meta es obtener aseveraciones cuantitativas respecto a los parámetros, como las regiones de confianza aproximadas (Díaz-Francés y Sprott 2000); las cuales pueden construirse con el método de Wald (Meeker y Escobar, 1998). Para la estimación con máxima verosimilitud, desde datos censurados, el algoritmo EM ha sido una herramienta muy útil (Flury y Zoppè, 2001), implementándose en paquetes estadísticos como *R* (2006). En este trabajo se propone una alternativa paramétrica para comparar concentraciones medias de contaminantes de muestras lognormales, haciendo una reparametrización a través de un modelo de regresión de variables indicadoras, en presencia de datos no detectados y covariables; estableciendo un criterio de comparación mediante regiones de confianza. Una modificación del algoritmo EM es usada para ajustar los modelos, manejar los datos no detectados y los errores, obteniéndose regiones aproximadas del 95% de confianza. Se desarrollan ejemplos con datos de estudios ambientales, ilustrando su aplicación al comparar poblaciones independientes y muestras con datos agrupados.

2. Metodología

Máxima verosimilitud provee métodos versátiles para ajustar modelos y pueden ser aplicados con datos censurados; el ajuste considera combinaciones de parámetros-modelos para la cual la probabilidad de los datos sea grande. También comprende el ajuste con análisis de regresión. Estos métodos, en muestras grandes, proporcionan estimadores mas exactos, y tienen buen desempeño con muestras moderadas.

Función de Verosimilitud

La función de verosimilitud es igual o proporcional a la probabilidad conjunta de los datos. Para un conjunto de datos y modelo especificado $F(y; \theta)$, la verosimilitud es una función de

los parámetros desconocidos θ . Para una muestra censurada por la izquierda, de n observaciones independientes, la verosimilitud se define como:

$$L(\theta) = C \prod_{i=1}^n L_i(\theta; \text{datos}_i) = C \prod_{i=1}^n [f(y_i; \theta)]^{\delta_i} [F(y_i; \theta)]^{1-\delta_i} \quad (1)$$

donde $L_i(\theta; \text{datos}_i)$ es la probabilidad de la observación i , datos_i son los datos para la observación i , para una observación exacta $\delta_i = 1$ y $\delta_i = 0$ para un dato no detectado; C es una constantes que no depende de θ . El valor de θ que maximiza $L(\theta)$ provee el estimador de máxima verosimilitud (EMV) para el vector θ , denotándose por $\hat{\theta}$.

El Algoritmo EM

El algoritmo EM es una herramienta poderosa para calcular estimadores de máxima verosimilitud con datos incompletos o censurados (Flury y Zoppè 2001). Sean y los datos observados y x los datos desconocidos. Entonces, el algoritmo EM es un proceso iterativo para calcular $\hat{\theta}$ asociado a la verosimilitud de los datos $L(y|\theta)$. Inicia con una primera aproximación $\theta^{(0)}$ y la iteración i -ésima consta de dos etapas:

Etapla E (Esperanza). Calcular $Q(\theta|\theta^{(i-1)}) = E(\log L(x|\theta)|y, \theta^{(i-1)})$.

Etapla M (Maximización). Obtener $\theta^{(i)}$ que maximiza en θ a $Q(\theta|\theta^{(i-1)})$.

El criterio de convergencia se basa en la norma entre $Q(\theta^{(i-1)}|\theta^{(i-1)})$ y $Q(\theta^{(i)}|\theta^{(i-1)})$. El algoritmo EM se simplifica con muestras aleatorias de la familia exponencial.

Regiones e Intervalos De Confianza Aproximados

La aproximación normal para la distribución de estimadores de máxima verosimilitud puede ser usada para calcular regiones de confianza aproximados para θ . Esto se conoce como Método de Wald, también conocido como el método de aproximación normal (Meeker y Escobar, 1998). En particular, una región aproximada del $100(1-\alpha)\%$ de confianza para θ es el conjunto de los valores de θ en el elipsoide

$$W(\theta) = (\hat{\theta} - \theta)' (\hat{\Sigma}_{\hat{\theta}})^{-1} (\hat{\theta} - \theta) \leq \chi_{(1-\alpha; r)}^2, \quad (2)$$

donde r es la dimensión de θ , $\hat{\Sigma}_{\hat{\theta}}$ es la matriz de varianzas y covarianzas estimada, $\chi_{(1-\alpha; r)}^2$ es el $1-\alpha$ cuantil de la variable aleatoria χ^2 con r grados de libertad.

Como se muestra en (Meeker y Escobar, 1995), esta región de confianza puede ser vista como una aproximación cuadrática para la log-verosimilitud perfil de θ in $\hat{\theta}$.

Entonces, un intervalo de confianza aproximado del $100(1-\alpha)\%$ se obtiene de $[\hat{\theta}, \tilde{\theta}] = \hat{\theta} \pm z_{(1-\alpha/2)} \hat{se}_{\hat{\theta}}$, donde $\hat{se}_{\hat{\theta}}$ es el error estándar estimado, $z_{(1-\alpha/2)}$ es el $1-\alpha/2$ cuantil de la distribución normal estándar.

3. Modelos Estadísticos

Distribución de Probabilidad Lognormal

Investigadores en ciencias ambientales han reportado que las concentraciones de contaminantes en aire y suelo, y de residuos de metales en ríos, tienen distribución lognormal (Gilbert, 1987; Ott 1995). Si y es una variable aleatoria lognormal de dos parámetros, tiene función de densidad de probabilidad,

$$f(y; \mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2}\right], \quad 0 < y < \infty, -\infty < \mu < \infty, \sigma > 0. \quad (3)$$

La mediana de y , $M = e^\mu$, depende solo de μ ; en contraste, su media $E = e^{\mu + \sigma^2/2}$, depende de μ y σ , por lo que la inferencia sobre la media requiere de un análisis simultáneo para ambos parámetros.

Comparación de Medias de dos Muestras Lognormales Independientes

La función log-cuantil del modelo de regresión de log-localización-escala para la distribución lognormal (Meeker y Escobar, 1995), involucrando la variable dummy x , para la comparar dos muestras, esta dada por,

$$\ln[y_p(x)] = \mu(x) + \Phi^{-1}(p)\sigma = \beta_0 + \beta_1 x + \Phi^{-1}(p)\sigma(x) \quad (4)$$

Donde Φ es la función de distribución de la normal estándar; respecto a la variable indicadora, $x = 0$ para una muestra y $x = 1$ para la otra.

La comparación de medias se realiza mediante variables indicadoras en los modelos de regresión para μ y σ , reescribiendo (4) con la reparametrización $\mu_i(x) = \beta_0 + \beta_1 x$, $\ln(\sigma_i) = \ln[\sigma(x)] = \gamma_0 + \gamma_1 x$ y optimizando su verosimilitud. Entonces, $\mu_1 = \beta_0$, $\ln(\sigma_1) = \gamma_0$ para la primera muestra y $\mu_2 = \beta_0 + \beta_1$, $\ln(\sigma_2) = \gamma_0 + \gamma_1$ para la segunda. Una región de confianza permitirá, simultáneamente, analizar si β_1 y γ_1 son ceros, y aseverar respecto a la diferencia significativa entre las medias poblacionales. Entonces, ambas concentraciones medias son iguales si $W(0) = \hat{\theta}' \hat{\Sigma}_{\hat{\theta}}^{-1} \hat{\theta} \leq \chi^2_{(1-\alpha; 2)}$.

Comparación de Medias de Dos Poblaciones Usando Modelos Mixtos

Los modelos lineales de efectos mixtos son usados para describir la relación entre una variable respuesta, y covariables en los datos agrupados de acuerdo a factores de clasificación; asociando efectos aleatorios comunes a las observaciones del mismo nivel o factor de clasificación (Pinheiro y Bates 2004). Estos modelos se definen así,

$$Y = X\beta + Zb + \varepsilon; \quad b \sim N(0, \Psi), \quad \varepsilon \sim N(0, \sigma^2 I), \quad (5)$$

β es el vector de efectos fijos y b de efectos aleatorios; X , Z son matrices de diseño, ε es el vector de errores intra-grupo. Se asumen b y ε independientes. La estimaciones se obtienen usando algoritmo EM y el método general por máxima verosimilitud (Pinheiro y Bates 2004).. La comparación de medias, se realiza a través del modelo,

$$\ln(y) = \mu(x) + \alpha_0 t + b_0 = \beta_0 + \beta_1(x) + \alpha_0 t + b_0 \quad (6)$$

β_0 , β_1 , α_0 , son los parámetros fijos y b_0 el parámetro aleatorio debido al agrupamiento de los datos. Asumiendo homogeneidad de σ , $\mu_1 = \beta_0 + \alpha_0 t$ y $\mu_2 = \beta_0 + \beta_1 + \alpha_0 t$. Así, la inferencia sobre la igualdad de medias dependerá solo del intervalo para β_1 .

4. Ejemplos de Aplicación

Comparación de Medias de Dos Poblacionales Independientes

En algunos casos, un grupo tratamiento es comparado con uno control, donde el primero representa las condiciones donde se sospecha que las concentraciones de contaminantes son altas. El interés es propiamente saber si los niveles de contaminantes en los grupos son iguales o diferentes.

Cuadro3. Concentraciones de Tolueno para dos muestras en un periodo 5 meses, US EPA (1992).

| Mes | Sitio en Riesgo | | Sitio Control | | |
|-----|-----------------|--------|---------------|--------|--------|
| | Pozo 1 | Pozo 2 | Pozo 1 | Pozo 2 | Pozo 3 |
| 1 | < 5 | < 5 | < 5 | < 5 | < 5 |
| 2 | 7.5 | < 5 | 12.5 | 13.7 | 20.1 |
| 3 | < 5 | < 5 | 8.0 | 15.3 | 35.0 |
| 4 | < 5 | < 5 | < 5 | 20.2 | 28.2 |
| 5 | 6.4 | < 5 | 11.2 | 25.1 | 19.0 |

Usando lme de R y agrupando de acuerdo a los pozos en cada sitio, se ajusta el modelo descrito en 3.3, obteniéndose las estimaciones para los parámetros, Cuadro 4.

Cuadro 4. Datos de Concentraciones de Tolueno (ppm) para dos muestras en un periodo 5 meses

| Parámetro | EMV | Error Estándar | IC Aprox. del 95% |
|------------|--------|----------------|-------------------|
| β_0 | 1.792 | 0.348 | 1.110, 2.474 |
| β_1 | -1.353 | 0.284 | -1.911, -0.796 |
| α_0 | 0.210 | 0.098 | 0.019, 0.401 |

| μ_1 | μ_2 | σ_b | σ_ε | $\log L(\beta, \gamma, \delta)$ |
|---------|---------|------------|----------------------|---------------------------------|
| 0.439 | 1.792 | 0.104 | 0.690 | -27.664 |

Note que $\hat{\beta}_1 = -1.353$, con un intervalo que no contiene el cero, $(-1.911, -0.796)$, de modo que sí existe evidencia de que ambas concentraciones medias son distintas. Entonces el sitio control tiene una concentración mayor de tolueno, por lo que no existe evidencia de contaminación en el sitio en riesgo. De la inferencia sobre α_0 se observa un ascenso sobre el tiempo en las concentraciones. Y el error estimado del efecto aleatorio debido al agrupamiento σ_b , resultó relativamente menor, 0.690.

4. Conclusiones

La metodología presentada, resultó versátil y simple para comparar poblaciones lognormales mediante estructuras de regresión. Se mostró la comparación de dos poblaciones independientes mediante una reparametrización en el modelo. El criterio de comparación resultó eficiente al observar un intervalo de confianza aproximado y el estadístico de Wald. El enfoque desde los modelos lineales mixtos presentado extiende el método de comparación con efectos aleatorias debido a datos agrupados, manteniéndose el criterio de comparación basado en los intervalos de confianza. El algoritmo EM mejoró sustancialmente la optimización de las funciones de verosimilitud. En los ejemplos mostrados se compararon solo dos poblaciones, sin embargo, el procedimiento puede extenderse a tres o más poblaciones lognormales.

REFERENCIAS

- Díaz-Francés, E., & Sprott, D.A. (2000). The use of the likelihood function in the analysis of environmental data. *Environmetrics 11*, 75–79.
- El-Shaarawi, A. H., & Viveros, R. (1997). Inferences about the mean in log-regression with environmental applications. *Environmetrics 8*, 569–582.
- EPA (1992). *Statistical Training Course for Ground-Water Monitoring Data Analysis*. EPA530-R-93-003. Office of Solid Waste. U.S. Environmental Protection Agency, Washington, DC.
- Flury, B., & Zoppè, A. (2001). Exercise in EM. *The Am. Statist. 54*, pp. 207 - 209.
- Gilbert, R.O. (1987). *Statistical Methods for Environmental Pollution Monitoring*. New York: Wiley.
- Helsel, D. R. (2005). *Nondetects And Data Analysis*. New York: Wiley.
- Meeker, W. Q., & Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician 49*, 48-53.
- Meeker, W. Q., & Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: Wiley
- Millard, S. P., & Deverel, S. J. (1988). Nonparametric statistical methods for comparing two sites. *Water Resources Research 24*, 2087–2098.
- Millard, S. P., & Neerchal, N. K. (2001). *Environmental Statistics with S-PLUS*. FL: CRC Press
- Ott, W. R. (1995). *Environmental Statistics and Data Analysis*, FL: CRC Press.
- Pinheiro, J.C., & Bates, D.M. (2004). *Mixed-Effects Models in S and S-PLUS*, New York: Springer
- R Core Team. (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.

Modelos de Regresión Múltiples para Explorar el Contexto Paisajístico sobre la Biodiversidad en un Agropaisaje Mesoamericano

S. Vílchez¹, CA. Harvey²; F. Casanoves³, B. Finegan⁴, J. Saenz⁵, D. Sánchez⁶, A. Medina⁷, Hernández⁸

¹Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). svilchez@catie.ac.cr

²Conservacion Internacional (CI).

³Centro Agronómico Tropical de Investigación y Enseñanza (CATIE).

⁴Centro Agronómico Tropical de Investigación y Enseñanza (CATIE).

⁵**Instituto Internacional en Conservación y Manejo de Vida Silvestre Universidad Nacional de Costa Rica.**

⁶Centro Agronómico Tropical de Investigación y Enseñanza (CATIE).

⁷**Fundación Amigos del Río San Juan (FUNDAR).**

⁸**Museo Entomológico de León.**

Resumen

Se evaluó el contexto paisajístico (composición y estructura del hábitat y del agropaisaje) sobre dos grupos faunísticos con capacidades de dispersión distintas (murciélagos y escarabajos coprófagos) en un agropaisaje en Matiguas, Nicaragua. El muestreo de la diversidad se realizó en seis diferentes tipos de hábitat: bosques secundarios, bosques ribereños, charrales, cercas vivas, potreros con alta cobertura arbórea y potreros con baja cobertura arbórea, con ocho parcelas por hábitat. Se caracterizó la comunidad de murciélagos capturados mediante redes de niebla y la comunidad de escarabajos coprófagos capturados con trampas *pitfall*. Además, se realizó el muestreo de la vegetación para caracterizar cada tipo de hábitat. Se calcularon métricas de la composición y arreglo espacial del paisaje a distintas escalas, para determinar la escala a cual cada grupo faunístico respondió al paisaje. Se establecieron las relaciones entre las características del hábitat y del paisaje con la diversidad de cada grupo faunístico a través de regresiones múltiples. La riqueza, abundancia y diversidad de murciélagos y escarabajos respondieron a distintas características del hábitat y del agropaisaje en distintas escalas. Las características del hábitat a la que respondieron los distintos atributos de la comunidad de murciélagos fueron: riqueza de árboles, área basal, diversidad de arbustos, riqueza y abundancia de flores; en cambio, las variables de paisaje fueron: distancia a los potreros con baja cobertura arbórea, porcentaje de bosque secundario y de cercas vivas, densidad de parches de charrales, de cercas vivas, de potreros con alta y baja cobertura arbórea, densidad de la riqueza de hábitat y diversidad y equidad de hábitat. Los atributos de la comunidad de escarabajos respondieron a variables de hábitat como altura y diámetro de los árboles y a variables de paisaje como porcentaje de bosque secundario, cercas vivas, potreros con alta y baja cobertura arbórea, densidad de parches de bosques ribereños, bosques secundarios y densidad de potreros con alta cobertura arbórea y baja cobertura arbórea. Ningún atributo de la comunidad de murciélagos y escarabajos respondió por igual a todas las variables de hábitat y de paisaje. La comunidad de murciélagos respondió a variables del agropaisaje medidas a distintas escalas; para algunos atributos de la comunidad las variables parecen influir a escalas locales (100 y 250 m) en ciertas variables de hábitat, mientras que para otros atributos como riqueza y abundancia de murciélagos frugívoros fueron más

importantes a escalas más grandes (1500 a 2500 m). Los escarabajos respondieron mayormente a características del agropaisaje medidas a escalas locales (100 a 500 m). Para mantener biodiversidad de escarabajos y murciélagos en el agropaisaje se debe aumentar el porcentaje de bosque secundario, bosque ribereño y disminuir el aislamiento dentro de estos a través de aumentar la cobertura arbórea sea a través de cercas vivas o árboles en potreros.

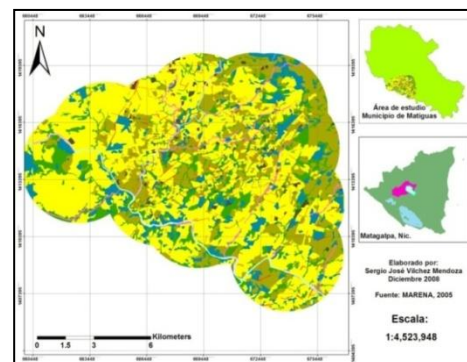
Palabras clave: biodiversidad, escarabajos coprófagos, escala, hábitat, murciélagos.

1. Introducción

Estudios recientes demuestran que el mosaico de hábitats que compone a un paisaje agropecuario puede albergar algunas especies prioritarias para la conservación, reducir el efecto de la fragmentación (al conectar parches de bosques dentro del paisaje) y mantener una alta biodiversidad producto de las diferencias en la composición y estructura de los hábitats que lo conforman. Sin embargo, hasta la fecha no se ha podido establecer cómo el arreglo espacial o la composición del agropaisaje influye en la diversidad de los distintos organismos ahí presentes, ni existen muchos estudios que demuestren la importancia relativa de diferentes hábitats dentro de los agropaisajes para diferentes grupos taxonómicos. Entender la importancia de estos sistemas en la manutención de la biodiversidad en la región mesoamericana es relevante, puesto que los paisajes boscosos continuamente se transforman en paisajes dominados por pasturas y tierras agrícolas (Harvey et ál. 2005). En Nicaragua, por ejemplo, los bosques se transforman a un ritmo de 130141 ha por año (tasa de deforestación reportada hasta el año 2000) y se predice que el ritmo del avance de la frontera agrícola para el 2010 será de 150000 ha por año (INAFOR 2004). Además, la cobertura boscosa que se encuentra bajo protección en Nicaragua (16% del territorio nacional) está sumergida en una matriz de áreas agropecuarias, compuesta por diferentes sistemas de producción y remanentes de bosques, que cubren el 46% del territorio. Conservar la biodiversidad en estos complejos paisajes requiere comprender la respuesta que los distintos grupos faunísticos ahí presentes tienen a la diversidad de hábitats y su arreglo espacial en el paisaje. El propósito de este estudio fue explorar la respuesta de distintos grupos taxonómicos (murciélagos y escarabajos estercoleros) a la estructura y composición del hábitat en un agropaisaje en Matiguás, Nicaragua, buscando responder si éstos responden a las mismas variables o si difieren según sus requerimientos alimenticios o de hábitat.

Descripción del área de estudio

El estudio se realizó en la subcuenca del río Bul Bul del Municipio de Matiguás, Departamento de Matagalpa, Nicaragua, con un área aproximada de 9200 km² (**¡Error! No se encuentra el origen de la referencia.**). La principal actividad en la zona es la ganadería de doble propósito (Ruiz et ál. 2005). El agropaisaje tiene un área de 160.12 km² y se encuentra dominado por potreros de baja cobertura arbórea (Cuadro 1).



Cuadro 1. Estructura espacial del agropaisaje de Matiguás, Nicaragua, en base a una imagen Ikonos (2003).

| Tipo de cobertura de suelo | Área total (ha) | % de área total | No. de parches | Tamaño medio del parche (ha) | Desviación estándar |
|------------------------------------|-----------------|-----------------|----------------|------------------------------|---------------------|
| Bosques secundarios | 2034.20 | 12.70 | 342 | 5.95 | 12.73 |
| Bosques ribereños | 306.05 | 1.91 | 72 | 4.25 | 5.15 |
| Charrales | 1613.26 | 10.07 | 282 | 5.72 | 8.85 |
| Cercas vivas | 235.81 | 1.47 | 368 | 0.64 | 1.15 |
| Potreros de alta cobertura arbórea | 3371.34 | 21.05 | 341 | 9.89 | 40.93 |
| Potreros de baja cobertura arbórea | 8087.64 | 50.50 | 404 | 20.02 | 74.41 |
| Otros | 366.05 | 2.28 | 81 | 4.51 | 23.60 |
| Total | 16014.34 | 100 | 1890 | 8.65 | 40.54 |

2. Metodología

El estudio de la biodiversidad de murciélagos se realizó en 6 tipos de hábitats: bosques secundarios (BS), bosques ribereños (BR), charrales (CH), cercas vivas (CV), potreros de alta cobertura arbórea (PAC, cobertura de 16-25%) y potreros de baja cobertura arbórea (PBC, cobertura de 1-5%). A través de una imagen Ikonos (2002) se seleccionaron al azar, para cada hábitat, ocho parcelas de 1 ha, para un total de 48 parcelas, en las cuales se realizó el muestreo de los murciélagos y escarabajos, los primeros a través de su captura con redes de niebla (mayores detalles Medina et ál. 2007), y los escarabajos coprófagos con trampas *Pitfall*. Se calculó la riqueza y el índice de diversidad de Shannon para cada una de las 48 parcelas. Los escarabajos se separaron según sus hábitos coprófagos en cavadores y rodadores.

VARIABLES explicativas

Para determinar la estructura y composición del hábitat en las 48 parcelas se midieron distintas variables de vegetación en parcelas de 0.1 ha (20 m x 50 m) (mayores detalles Sánchez et ál. 2006). Para caracterizar la estructura y composición del paisaje se utilizó una imagen Ikonos (2003) con 1 m de resolución, y mediante la elaboración de un mapa se caracterizaron ocho tipos de hábitat. Se calcularon variables de paisaje (composición y estructura) con el programa *Fragstat* 3.3 alrededor de cada una de las 48 parcelas de muestreo dentro de círculos de 100, 250, 500, 1000, 1500, 2000, 2500 y 3000 m.

Análisis de los datos

Por la sospecha de existencia de autocorrelación espacial de las variables de respuesta y por la sobreposición de los círculos donde se midieron las variables de paisaje, y para evitar redundancia en la información proporcionada por las métricas, se realizaron análisis exploratorios de autocorrelación espacial y correlaciones de *Spearman*. Para determinar cuáles variables de la estructura y composición del hábitat y del agropaisaje determinan la diversidad de murciélagos se realizaron modelos de regresión múltiple, en donde las variables de estructura y composición del paisaje y de los hábitats fueron las variables

regresoras y los atributos de la comunidad de murciélagos y escarabajos fueron las variables de respuesta. La retención de variables se realizó con el método de selección *backward*. Los criterios de selección de los modelos fueron el R^2 ajustado, el error cuadrático medio de predicción (ECMP) y el cuadrado medio del error (CMerror).

3. Resultados

Las regresiones múltiples finales mostraron que la comunidad de murciélagos y escarabajos coprófagos respondieron tanto a variables de hábitat como a variables de composición y estructura del paisaje, pero las variables importantes difirieron para diferentes atributos de la comunidad de murciélagos y escarabajos. Además, estas respuestas se dieron en múltiples escalas, desde variables medidas a la escala de hábitat (< 100 m) hasta variables de composición del paisaje a una distancia de 3000 m (Cuadro 1, 2, 3).

Cuadro 2. Variables de paisaje y de hábitat que explicaron los patrones de la diversidad de murciélagos en los modelos de regresión múltiple; el porcentaje de la suma de cuadrados (%SC) indica el peso de cada variable en el modelo.

| Variables regresoras | Riqueza | | | H' | | |
|---|----------------------|--------------|---------|----------------------|--------------|---------|
| | R ² ajust | ECMP | CMerror | R ² ajust | ECMP | CMerror |
| | 0.50 | 6.55 | 4.67 | 0.57 | 0.1316 | 0.0847 |
| | Relación | % SC | P | Relación | % SC | P |
| Abundancia de flores | - | 4.91 | 0.0181 | | | |
| Área basal de los árboles | | | | - | 5.96 | 0.0134 |
| Distancia a PBC | | | | - | 7.98 | 0.0049 |
| Porcentaje de área de BS 1000m | + | 7.1 | 0.0051 | | | |
| Porcentaje de área de CV 500m | + | 24.66 | 0.0001 | | | |
| Densidad de parches de CH 1500m | - | 6.76 | 0.0061 | - | 20.74 | 0.0001 |
| Densidad de parches de PAC 1000m | | | | - | 6.7 | 0.0092 |
| Densidad de parches de PBC 100m | + | 16.71 | 0.0001 | | | |
| Densidad de parches de PBC 1000m | | | | + | 9.68 | 0.0021 |
| Riqueza de la densidad de parches a 1500m | + | 44.92 | 0.0001 | + | 37.75 | 0.0001 |
| Índice de equidad de hábitat a 1500m | | | | + | 20.31 | 0.0001 |

Cuadro 3. Variables de paisaje y de hábitat que explicaron los patrones de la diversidad de escarabajos coprófagos en los modelos de regresión múltiple.

| Variables regresoras | Riqueza | | | Índice de diversidad | | |
|---------------------------------|---------|-------|----------|----------------------|-------|---------|
| | R2ajust | ECMP | CMerror | R2ajust | ECMP | CMerror |
| | 0.26 | 11.3 | 9.61 | 0.17 | 0.57 | 0.48 |
| Relación | %SC | P | Relación | %SC | P | |
| Altura promedio de los árboles | | | | + | 8.82 | 0.046 |
| Diámetro altura del pecho | | | | - | 16.27 | 0.0079 |
| Porcentaje área de BS 500m | + | 16.06 | 0,0034 | | | |
| Densidad de parches de CV 2000m | + | 6.97 | 0,0416 | | | |
| Densidad de parches BR 500m | - | 3.46 | 0,1461 | | | |

Cuadro 4. Variables de paisaje y de hábitat que explicaron los patrones de la diversidad de grupos de escarabajos coprófagos en los modelos de regresión múltiple.

| Variables regresoras | Riqueza de cavadores | | | Riqueza de rodadores | | |
|--------------------------------|----------------------|-------|---------|----------------------|------|---------|
| | R2ajust | ECMP | CMerror | R2ajust | ECMP | CMerror |
| | 0.25 | 7.92 | 6.99 | 0.32 | 0.72 | 0.54 |
| | Relación | %SC | p-valor | Relación | %SC | p-valor |
| Altura promedio de los árboles | + | 10.69 | 0.0127 | I. | | |
| Distancia a CV | | | | + | 14 | 0.0032 |
| Porcentaje de área CH 100m | | | | + | 9.6 | 0.0133 |
| Porcentaje de área PBC 1500m | - | 15.19 | 0.0034 | | | |

4. Implicaciones para la Conservación

Para poder conservar y mantener mayor diversidad en estos hábitats se debe tomar en cuenta el arreglo espacial de las características que conforman el agropaisaje y la estructura de los hábitats. Aumentar el porcentaje de bosque secundario y de bosque ribereño, y disminuir el aislamiento dentro de estos, y reducir los potreros de baja cobertura arbórea, ayudaría a favorecer la conservación de la comunidad de escarabajos coprófagos en el agropaisaje. Al no haber una escala apropiada de la respuesta de la diversidad de murciélagos y escarabajos a la configuración del agropaisaje para la incidencia de toma de decisiones en el manejo y conservación, se puede decir que es necesario hacer un manejo a múltiples escalas, a nivel de fincas o de paisaje. Este estudio aporta al conocimiento sobre la importancia de la matriz agropecuaria para la conservación de la diversidad de murciélagos y escarabajos.

REFERENCIAS

- Harvey, C. A., Alpízar, A., Chacón, M y Madrigal, M. (2005). Assessing linkages between agriculture and biodiversity in Central America: Historical overview and future perspectives. Mesoamerican & Caribbean Region, Conservation Science Program. The Nature Conservancy (TNC), San José, Costa Rica.
- Instituto Nicaragüense de Fomento Municipal (INIFOM). (2001). Caracterizaciones Municipales. PNUD. Managua, NI. www.inifom.gob.ni
- Medina, A., Harvey, C. A., Sánchez, D., Vilchez, S. y Hernández, B. (2007). Bat diversity and movement in an agricultural landscape in Matiguás, Nicaragua. *Biotropica* **39**, 120–128.
- Sánchez, D., Harvey, C. A., Grijalva, A., Medina, A., Vilchez, S. y Hernández, B. (2006). Diversidad, composición y estructura de la vegetación en un agropaisaje ganadero en Matiguás, Nicaragua. *Biología Tropical*. **53**, 387-414.

AGRADECIMIENTOS

Al proyecto FRAGMENT por el apoyo y el permiso de los datos, a los productores de la zona por su apoyo en campo.

Diagnosis de Dengue: Una Aplicación de Regresión Logística Binaria.

Ana María Aguilera¹, Agustín Santiago²

¹Departamento de Estadística e Investigación Operativa. Facultad de Ciencias.
Universidad de Granada. España. aaguiler@ugr.es

²Universidad Autónoma de Guerrero. Unidad Académica de Matemáticas.
México. santmag@correo.ugr.es

1. Introducción.

El dengue es un síndrome febril agudo de origen viral, que representa un problema de salud pública en diferentes partes del mundo y representa una enfermedad endémica en América, Sudeste de Asia, Oeste del Pacífico, África y el este del Mediterráneo.

Considerada hoy como una enfermedad emergente que es causada por un virus que pertenece al grupo de los arbovirus que se transmite por la picadura de las hembras de mosquitos de la especie *Aedes aegypti*. La enfermedad está caracterizada por fiebre, mialgias, artralgias, exantema, leucopenia y linfadenopatía. Se han descrito tres cuadros clínicos causados por este virus, el dengue clásico (DC), el dengue hemorrágico (DH) y el síndrome de choque por dengue (SCD).

La vigilancia epidemiológica del dengue constituye un elemento esencial para la prevención y control de esta enfermedad. En México y específicamente en el estado de Guerrero, se han establecido los criterios para la toma, manejo y recepción de muestras para el diagnóstico por laboratorio del dengue, entre los que se incluye el diagnóstico serológico para la identificación de anticuerpos. Durante el proceso de diagnóstico del dengue, se incluye la valoración clínica en la unidad de salud y su clasificación como casos probables de Dengue Clásico (DC) o Dengue Hemorrágico (DH). Hecha esta clasificación se obtiene una muestra de suero que se envía al Laboratorio Estatal de Salud Pública (LESP) para su confirmación serológica, acompañada de un resumen clínico del caso y, a partir de la determinación de anticuerpos IgM contra el virus de dengue, se confirma o descarta el caso.

En 2006, el LESP recibió 16604 muestras para confirmación serológica, de los cuales resultaron positivos solo el 28.8 %. De estos casos, el índice de positividad para DC es del 23.4% y para DH es de 77.7%. Esta disparidad se acentúa en algunas regiones del estado, sugiriendo que existen diferencias en las habilidades diagnósticas de los expertos clínicos que realizan el primer contacto con los enfermos. De ahí la importancia de explorar algunos modelos que ayuden a tomar decisiones a los expertos clínicos y generar una herramienta probabilística que permita estimar con mayor certidumbre la presencia de la enfermedad. Un diagnóstico oportuno y con mayor certidumbre, significará un importante ahorro de recursos en el control epidemiológico de la enfermedad y su tratamiento médico.

Trabajaremos con una base de datos generados en un estudio previo sobre consenso¹, en el cual el interés principal se centró en evaluar las clasificaciones que hacen los expertos

¹ Análisis de la efectividad de la diagnosis del dengue en el estado de Guerrero, México 2007: hechos estadísticos y consenso. Sira Allende y otros

clínicos en el manejo de los pacientes sospechosos, empleando como herramienta de análisis índices del tipo kappa. Esta base está constituida por 2206 casos confirmados, en la cual se concentra también información sobre las distintas variables observadas en los pacientes sospechosos y que se describen en el documento denominado manual de dengue, que consiste principalmente en la sintomatología que se debe observar en un cuadro clínico de dengue².

2. Metodología.

Toda vez que nuestro objetivo consiste en proponer un modelo de diagnóstico de la enfermedad, debemos considerar como nuestra variable de interés, dada la sintomatología de un cuadro de dengue, si el paciente está afectado por dengue clásico o dengue hemorrágico. Esto se propone de esta forma porque es muy bajo el índice de positividad en el dengue clásico, lo cual no ocurre con el dengue hemorrágico, según los expertos clínicos por la mayor claridad en la manifestación de sus síntomas.

La variable respuesta será denotada por Y y tomará el valor $Y = 1$ cuando el paciente presente dengue clásico (DC) y $Y = 0$ cuando presente dengue hemorrágico (DH). Las variables explicativas que consideraremos son en su totalidad categóricas, la mayoría de ellas con solo dos categorías, estas son: Contacto con animales, Sexo, Enfermos similares, Transfusión sanguínea, Fiebre, Cefalea, Mialgias, Artralgias, Dolor retrocular, Exantema, Prurito, Vómito, Náuseas, Escalofrío, Fotofobia, Dolor abdominal, Diarrea, Conjuntivitis, Cogestión nasal, Tos, Faringitis, Rinitis, Hepatomegalia, Esplenomegalia, Escape de líquidos, Hemorragias, Petequias, Equimosis, Hematomas, Torniquete positivo, Ascitis, Derrame pleural, Gingival, Epistaxis, Hematemesis y Melena.

Dadas las características de las variables, el modelo que se propone es el de regresión logística binaria con variables categóricas y el método de selección que se empleará es el método *stepwise* de Wald incorporado en el paquete estadístico SPSS.

3. Ajuste del modelo de regresión logística.

El modelo inicial incluye 36 variables regresoras, pero utilizando el método de selección *stepwise* finalmente obtenemos 14 variables candidatas a entrar en el modelo, las cuales son: El, Da, He, Es, V, M, N, Ce, Ar, P, Co, S, Di y Cn.

El modelo puede escribirse de manera compacta dado que todas las variables explicativas son categóricas y binarias asociadas a los síntomas.

$$P[Y = 1|x_{i1}, \dots, x_{i14}] = \frac{e^{L_i}}{1 + e^{L_i}} = \frac{1}{1 + e^{-L_i}}$$

Donde

$$L_i = \beta_0 + \sum_{j=1}^{14} \beta_j x_{ij}, j = 1, \dots, 14; i = 1, \dots, n.$$

$\beta_0 = \text{constante}$ y β_i es el i -ésimo parámetro asociado a la i -ésima variable explicativa.

² Manual para la Vigilancia, Diagnóstico, Prevención y Control del Dengue. Secretaría de Salud. México. 2006.

Según las pruebas de razón de verosimilitudes y de Hosmer y Lemeshow, el modelo es bueno, toda vez que $-2 \log$ de la verosimilitud = 1769.576051368 y el p-valor > 0.05 para la prueba de Hosmer y Lemeshow en el último paso del método.

Los estimadores de máxima verosimilitud de los parámetros son:

$$\hat{\beta}_0 = -1.762, \hat{\beta}_1 = 0.274, \hat{\beta}_2 = -0.606, \hat{\beta}_3 = -1.011, \hat{\beta}_4 = 0.656, \hat{\beta}_5 = 0.391, \hat{\beta}_6 = 0.507, \hat{\beta}_7 = 0.624, \hat{\beta}_8 = 0.357, \hat{\beta}_9 = -0.402, \hat{\beta}_{10} = -0.360, \hat{\beta}_{11} = 1.386, \hat{\beta}_{12} = 0.976, \hat{\beta}_{13} = 0.721, \hat{\beta}_{14} = 1.633$$

Para contrastar la significación estadística de cada uno de los parámetros del modelo usamos el test estadístico Z de Wald, obtenido como el cociente entre el valor estimado del parámetro y su error estándar, pero como el resultado se ha generado en SPSS, trabajaremos con χ^2 , fijando un nivel de significación $\alpha = 0.05$ para rechazar la hipótesis de igualdad a cero de los parámetros cuando $\chi_0^2 \geq \chi_{1,\alpha}^2$ o cuando p-valor < 0.05 . Como consecuencia de esta prueba, son significativamente distintos de cero todos los parámetros considerados.

La relación entre dengue y cada una de las variables explicativas se estima a partir del cociente de ventajas a favor de padecer dengue clásico que viene dado por la exponencial del parámetro asociado a cada variable. En este trabajo se ha fijado como categoría de referencia la primera, de manera que para ver la relación que existe entre cada una de las variables explicativas y el padecimiento dengue debemos tomar en cuenta este hecho, además de que hemos supuesto un modelo de no interacción, lo cual supone que la interpretación dada para cada una de las variables es independiente de las otras.

En el caso del sexo el parámetro asociado es 0.274 de modo que su exponencial es 1.315. Esto significa que la ventaja a favor de padecer dengue clásico es 1.315 veces mayor para las mujeres que para los hombres incluidos en la población estudiada, es decir el segmento de la población constituido por mujeres tiene mayor probabilidad de presentar la enfermedad. Procediendo de forma análoga con el resto de los parámetros se tiene que haber padecido enfermedades similares, tener cefaleas, padecer conjuntivitis y tener congestión nasal son síntomas que llevan a una probabilidad alta de padecer dengue clásico, mientras que el padecimiento de Artralgia, vómito, náuseas, dolor abdominal, diarrea, escape de líquidos, hemorragias, Petequias y síntoma Melena están asociados con mayor probabilidad con el Dengue hemorrágico.

Para hacer una validación completa del modelo procederemos al análisis de los residuos. Sabemos que los residuos son significativos al nivel de confianza $\alpha = 0.05$, cuando sus valores ajustados son, en valor absoluto, mayores o iguales que 1.96. En este caso tenemos 132 de 2215 residuos estandarizados mayores que 1.96, que representan aproximadamente 6% de los datos. Dado que los valores de influencia son muy pequeños (el máximo toma el valor 0.0886 y el mínimo el valor 0.00139), estos residuos no influyen significativamente en la estimación de los parámetros del modelo.

Una vez estimado el modelo logit se puede usar como test de diagnóstico para clasificar la enfermedad en DC o DH en fallos o éxitos en función de los niveles de las variables. La tabla de clasificaciones considerando el modelo de no interacción con punto de corte 0.5 es

la que se muestra a continuación, de manera que a partir de esta se pueden interpretar los resultados

Tabla de clasificación(a)

| Observado | | Pronosticado | | |
|-------------------|----|--------------|-----|---------------------|
| | | Dengue | | Porcentaje correcto |
| | | DC | DH | |
| Dengue | DC | 1459 | 73 | 95.2 |
| | DH | 306 | 196 | 39 |
| Porcentaje global | | | | 81.4 |

La tasa de clasificaciones correctas puede ser considerada como buena, toda vez que estamos interesados, principalmente en la estimación de la probabilidad de padecer dengue clásico y, en este caso, la tasa de clasificaciones correctas es del 95.2%.

4. Conclusiones.

El modelo de regresión logística binaria es una buena herramienta para diagnosticar el dengue clásico en los pacientes considerados como sospechosos si es que se observan las catorce variables consideradas en el modelo. Para una correcta utilización en el diagnóstico, es importante que el experto clínico, que generalmente es el médico tratante, tenga claridad en lo que significa la ausencia o presencia de cierto síntoma en términos del modelo o en su defecto, ser experto en la identificación de los síntomas y estimar la probabilidad de que su paciente tenga dengue clásico o hemorrágico, a partir del modelo de probabilidad que se puede construir con el modelo logit obtenido en el presente trabajo.

La decisión que el experto tome respecto a la clasificación en uno u otro grupo debe hacerse a partir de las siguientes consideraciones que se derivan de la forma en cómo se ha generado el modelo. Primero, el software, a partir de los datos de cada paciente en las catorce variables explicativas, calcula una probabilidad para cada uno de ellos. Dado que se trata de una probabilidad, sus valores oscilarán entre 0 y 1, de modo que cuando más tienda a uno, más probable será que el paciente tenga dengue tipo clásico, al revés cuanto más tienda a 0. El software clasificará de este modo a los pacientes en enfermos de DC o DH a partir de un punto de corte que por defecto está situado en 0.5. La proporción de individuos correctamente clasificados en este modelo, como se ha mencionado antes, es uno de los indicadores de eficacia y del ajuste del mismo. Recordemos que para el modelo ajustado la tasa global de clasificaciones correctas es del 84.1% y para el caso particular del dengue clásico es del 95.2%.

La estimación de probabilidades puede hacerse en una hoja de cálculo tipo Excel o en una agenda de bolsillo tipo palm, creando la fórmula, sustituyendo los valores estimados de los parámetros en el modelo y los valores de las variables explicativas observadas en el paciente. El valor de salida será una probabilidad que deberá interpretarse de acuerdo al criterio descrito en esta sección.

REFERENCIAS

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. Wiley.
 Agresti, A. (2002). *Categorical Data Analysis (2ª edición)*. Wiley.

- Aguilera del Pino, A.M. (2001). Tablas de Contingencia Bidimensionales. Colección Cuadernos de Estadística (15). La muralla.
- Aguilera del Pino, A.M. (2006). Modelización de tablas de contingencia multidimensionales (Colección Cuadernos de Estadística, 33). La Muralla.
- Aguilera del Pino, A.M. (2008). Modelos de respuesta categórica y aplicaciones biosanitarias. Apuntes del curso
- Christensen, R. (1997). Log-Linear Models and Logistic Regression. Springer.
- Hosmer, D.W. y Lemeshow, S. (1989). Applied Logistic regression. Wiley.
- Kleinbaum, D.G. (1994). Logistic Regression. A Self-Learning Text. Springer.
- McCullagh, P. y Nelder, J. (1983). Generalized linear models. Chapman and Hall.
- Power, D.A. y Xie, Y. (2000). Statistical Methods for Categorical Data Analysis. Academia Press.
- Ruiz-Maya, L., Martín Pliego, F.J., Montero, J.M. y Uriz Tomé, P. (1995). Análisis Estadístico de Encuestas: Datos Cualitativos. Ed. AC.
- Selvin, S. (1996). Statistical Analysis of Epidemiological Data. Oxford University Press.
- Visauta Vinacua, B. (2003). Análisis estadístico con SPSS para Windows, Vol. II, Estadística multivariante. Mc Graw Hill.

Algunos Modelos Epidemiológicos Estocásticos

Cruz-Kuri, Luis¹, Sosa-Galindo, Ismael², García-Banda, Agustín J.³

¹Instituto de Ciencias Básicas. UV. kruz1111@yahoo.com.mx

²Facultad de Ciencias Administrativas y Sociales. U.V. isoga77@yahoo.com.mx

³Facultad de Ciencias Administrativas y Sociales, U.V. jaimegarciabanda@yahoo.com

1. Introducción

Los recientes brotes de enfermedades en animales, tales como los de la fiebre porcina clásica, la influenza aviar, y actualmente la influenza AH1N1, así como las constantes amenazas de nuevas infecciones pandémicas, hacen evidente la importancia del entendimiento y una capacidad de predicción de la dinámica de la propagación de una infección.

Para tales propósitos, existen una variedad de modelos, dentro de los cuales se tienen los matemáticos, los que hacen uso intensivo de la computadora y los correspondientes a procesos estocásticos. Cada uno de ellos tiene sus ventajas, alcances y limitaciones.

Para el presente trabajo se hablará de algunos modelos estocásticos, tales como los procesos de Galton-Watson, tanto para un solo tipo de estado individual como para muchos tipos. Para propósitos de estudios epidemiológicos, uno puede estar interesado principalmente en el desarrollo en tiempo real del número de individuos contagiados. Los procesos de Galton-Watson, aunque permiten estudiar el desarrollo de la epidemia, tienen ciertas limitaciones, las cuales se pueden evitar mediante la utilización de modelos más generales tales como los así denominados *procesos de ramificación*.

En lo que sigue se dará una breve descripción de los procesos de referencia y se presentarán algunos comentarios sobre resultados de aproximación del desarrollo de una epidemia por medio de los procesos de ramificación.

2. Nociones Básicas

Hace aproximadamente 75 años, W. O. Kermack y A. G. McKendrick publicaron una serie de artículos titulados “Contribuciones a la Teoría Matemática de las Epidemias”. Esos artículos se consideran con frecuencia como la base de desarrollos subsecuentes en la modelación matemática de la propagación de enfermedades infecciosas. En esas publicaciones, se consideran modelos epidemiológicos básicos de un solo tipo SIR (susceptible \rightarrow infeccioso \rightarrow removido). Las generalizaciones a modelos con más de un tipo son directas. En el modelo con un solo tipo se supone que la población es cerrada y de tamaño N , de tal forma que los nacimientos, las defunciones y las migraciones se ignoran. En ese modelo se supone que todos los pares de individuos tienen contactos a una tasa fija β/N . si un individuo infeccioso tiene contacto con un individuo susceptible, el individuo

susceptible se vuelve infeccioso. Se supone que la tasa de recuperación para los individuos infecciosos es una constante α .

Denótese con $S(t)$ al número de individuos susceptibles en el momento t y con $s(t)$ a la fracción $S(t)/N$ de la población que es susceptible en ese momento t . De manera análoga, se definen $I(t)$, $i(t)$, $R(t)$ y $r(t)$. Sean $i(0) = i_0$ y $s_0 = 1 - i_0$. Ahora se introduce el concepto de una tasa básica de reproducción (o número básico de reproducción) R_0 . La tasa básica de reproducción se define como el número de infecciones secundarias generadas por un individuo inicial infeccioso, en una población muy grande, donde $s(0)$ es aproximadamente igual a 1. En ese modelo, un individuo inicialmente infectado infectará en el promedio aproximadamente a otros β/α individuos. Así que, R_0 es aproximadamente igual a β/α .

Siguen algunas observaciones. *Primera.* Puede verse a partir de las ecuaciones diferenciales que modelan de manera determinista al fenómeno, que, si se introduce un número pequeño de individuos infecciosos (pequeño relativo al tamaño total de la población) el número de infecciosos inicialmente crecerá si $\beta - \alpha > 0$; es decir, si $R_0 > 1$. Por otra parte, si $R_0 \leq 1$, la epidemia se extinguirá rápidamente. Esta es la razón por la cual R_0 es una cantidad muy importante en epidemiología: constituye un parámetro de umbral, análogo al número promedio de descendientes en un proceso de ramificación. *Segunda.* La noción de la tasa básica de reproducción R_0 es muy importante para propósitos de control de la epidemia. *Tercera.* Al vacunar a una fracción menor que $1 - (1/R_0)$ de una población bien mezclada al principio, mediante una vacuna perfecta, puede conseguir que la tasa de reproducción en la población vacunada sea menor que 1 y de esta manera evitar que ocurra un brote mayor.

La duración de una epidemia no puede ser estudiada de manera realista mediante un modelo determinista, porque al principio y al final de una epidemia, las fluctuaciones aleatorias juegan un papel muy importante. Para algunas infecciones no es suficiente utilizar tres estados de infección (susceptible, infeccioso, y removido) sino que también se requiere tomar un estado de latencia o más de un estado infeccioso en consideración.

3. Modelos Epidemiológicos Estocásticos

3.1 El modelo de Reed-Frost

El modelo más básico es el así llamado “modelo de Reed-Frost”. Este modelo se basa en las “generaciones”. Supóngase que la población es cerrada, esto es, que durante la epidemia no se introducen nuevos individuos a la población y que no hay muertes. La probabilidad de que un individuo susceptible se vuelva infeccioso en la generación $k+1$ depende únicamente del número de infecciosos en la generación k . Se supone que, para cada individuo susceptible, la probabilidad de escapar a la infección de un individuo infeccioso específico es q . Sujeto a la condición de que el número de infecciosos de la generación k sea una cantidad dada, los eventos de que los distintos individuos susceptibles se escapen de ser infectados en la generación $k+1$ son todos independientes. Esto nos remite al modelo de la distribución Binomial con parámetros k , $\pi = 1 - q^L$, $1 - \pi = q^L$, condicionado a $S(j) = k$, $I(j) = L$. [Nota. $S(j+1) - S(j) = I(j+1)$].

3.2 El Modelo Epidemiológico Estocástico Estándar SIR

Para el modelo estocástico SIR se tienen las siguientes suposiciones:

1. La población es cerrada, así que no se consideran nacimientos, defunciones y migraciones en el modelo.
2. Inicialmente hay m individuos infecciosos y $N-m$ individuos susceptibles.
3. La población es homogénea y se mezcla de manera aleatoria; es decir, la probabilidad de que exista contacto entre dos individuos no depende de cuales individuos se consideren y además se supone que todos los individuos tienen las mismas características.
4. En cada contacto entre un individuo infeccioso y un individuo susceptible se transmite la infección.
5. Cada individuo se pone en contacto con otro individuo dado en instantes de tiempo que obedecen un proceso de Poisson con parámetro β/N . Abusando algo sobre la terminología a β se le puede llamar la *tasa de infección* o *tasa de contacto*.
6. Un individuo se dice que pertenece a la categoría R (de los casos removidos) si su periodo como infeccioso ha terminado.
7. Todos los individuos se distribuyen de manera independiente e idéntica en lo correspondiente a los periodos en que son infecciosos y la variable aleatoria correspondiente tiene un valor esperado τ con variancia finita σ^2 .

El proceso así descrito se denota con $E_{N-m,m}(\beta, \tau)$. De ahora en adelante, cuando se haga referencia a la palabra “tasa” en un marco estocástico, se entenderá la densidad de un proceso de Poisson unidimensional.

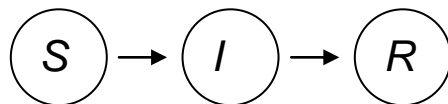


Fig. 1. Diagrama de flujo del modelo SIR.

Nota: S=susceptibles, I=infecciosos, R=removidos.

4. Procesos de Ramificación

Los procesos de ramificación inicialmente se desarrollaron para describir las probabilidades de extinción de los apellidos de familias prominentes y pueden considerarse como un estudio matemático de árboles genealógicos.

El proceso básico de ramificación (Galton-Watson) *a grosso modo* puede describirse como sigue: hay algunos ancestros (género masculino) cada uno de los cuales tiene un número aleatorio de hijos. Los números de hijos de los distintos ancestros se distribuyen de manera idéntica e independiente. Los hijos de los ancestros, llegado el momento, tienen hijos por si mismos, y los números de estos hijos quedan otra vez distribuidos de manera independiente

e idéntica como los descendientes de un ancestro. De esta manera, se puede definir un modelo para el árbol genealógico completo. Todos los individuos tienen de manera independiente un número de hijos, cada uno de estos obedeciendo una misma distribución de probabilidad.

El modelo puede enriquecerse considerando el árbol genealógico en tiempo real. En tal modelo extendido, no solamente el número de generación de un individuo es importante, sino también los tiempos de nacimiento y muerte. Tal modelo es útil y aun necesario si, por ejemplo, uno quiere hacer predicciones acerca de el número de gentes con el apellido, digamos, Pérez para el año 2100 (claro está, bajo las suposiciones de que la esperanza de vida, la distribución del número de hijos y la edad en que cada persona tiene a sus hijos no cambie entre ahora y el año 2100).

La relación entre la epidemia para una población grande, cuyos individuos se mezclan de manera aleatoria, y los árboles genealógicos, puede explicarse de manera intuitiva como sigue. Los individuos que introducen una infección en una población se ven ahora como los “ancestros”. Durante sus periodos como infecciosos, estos se ponen en contacto con un número aleatorio de individuos escogidos de manera uniforme en la población e infectan a esos individuos (si estos todavía son susceptibles). Los individuos recientemente infectados pueden verse como los individuos de la primera generación. Durante su periodo como infecciosos, los individuos de la primera generación se ponen en contacto con otros individuos de acuerdo a la misma ley que siguieron los ancestros.

Si la población es muy grande, la probabilidad de que un individuo infeccioso se ponga en contacto con un individuo que ya ha sido infectado durante la primera etapa de la epidemia (la cual puede tener una duración bastante grande) es muy pequeña, y, por consiguiente, el progreso de la epidemia puede describirse por medio de un proceso de ramificación.

En la siguiente sección se define de manera rigurosa el así llamado proceso de Galton-Watson y se presentan algunos resultados básicos de tal proceso. Además se describen algunas extensiones tales como los procesos multi-tipo de Galton-Watson. Para las demostraciones de los resultados y para una teoría adicional de procesos de ramificación uno puede consultar los tratados clásicos correspondientes. Por ejemplo, Jagers, P. (1975). *Branching processes with biological application*. Londres: John Wiley & Sons.

5. Definición del Proceso de Galton-Watson

En esta sección se consideran procesos de ramificación vistos desde la perspectiva de generaciones, a saber, los así llamados procesos de Galton-Watson. Sin pérdida de generalidad se puede suponer que solo hay un ancestro.

Un proceso de Galton-Watson es un proceso que genera un árbol genealógico, donde cada individuo tiene un número finito (no necesariamente acotado) de hijas. A ese ancestro se le etiquetará como a . La k_n -ésima hija de la k_{n-1} -ésima hija... de la k_1 -ésima hija del ancestro se etiquetará con el vector (a, k_1, \dots, k_n) . Si un individuo puede producir nacimientos de más de una hija al mismo tiempo, se puede utilizar un orden aleatorio para etiquetar a las hijas.

Al conjunto de todas las etiquetas posibles se le denotará con K . De ahora en adelante, se entenderá al “individuo x ” como aquel individuo que tiene una etiqueta (a,x) perteneciente a K y donde el “individuo a ” es el ancestro.

Nótese que hay demasiadas etiquetas, y en realidad, si el ancestro tiene solo una hija entonces no hay individuos con etiquetas de la forma $(a,2,k_2,\dots, k_n)$ con k_i en K y n número natural.

6. La Función Generadora y la Probabilidad de Extinción

Para una variable aleatoria discreta R se define su función generadora $f_R(s)$ por medio de

$$f_R(s) = E(s^R) = \sum P(R=k)s^k,$$

donde la sumatoria recorre los índices k desde 0 en adelante. Para todos los números reales s para los cuales esta serie sea convergente. Se puede asegurar que, si $s \leq 1$, entonces la serie es convergente. La función generadora tiene propiedades útiles, dentro de las cuales se encuentran la obtención de los momentos de la variable aleatoria mediante el cálculo de las derivadas de $f_R(s)$, de los órdenes correspondientes a los momentos requeridos. Asimismo, con la función generadora se puede establecer una ecuación funcional que permite calcular la probabilidad de extinción del proceso (de la epidemia en nuestro caso). Por consideraciones de espacio, no se presentan los detalles, aunque es pertinente mencionar que el tratamiento es estándar.

7. Comentarios Adicionales

La duración de una epidemia no puede ser estudiada de manera realista mediante un modelo determinista porque al principio y al final de una epidemia las fluctuaciones aleatorias juegan un papel muy importante. Para algunas infecciones no es suficiente utilizar tres estados de infección (susceptible, infeccioso y removido) sino que también se requiere tomar un estado de latencia o más de un estado infeccioso en consideración.

REFERENCIAS

- [1] Bailey, Norman T. J. (1964) *The Elements of Stochastic Processes with Applications to the Natural Sciences*. Wiley. New York.
- [2] Harris, Theodore E. (1989) *The Theory of Branching Processes*. Dover Books on Advanced Mathematics.
- [3] Kermack, W. O. and McKendrick, A.G. (1927), Contributions to the mathematical theory of epidemics, *Bulletin of Mathematical Biology* 53 (1/2) 33-118.
- [4] Sheldon, M. R. (1996) *Stochastic Processes*. Wiley. New York.
- [5] Trapman, Jan Pieter (2006) *On stochastic models for the spread of infections*. Impreso por PrintPartners Ipskamp, Enschede.

Análisis Multinivel: Una Generalización del Análisis de Covarianza de Efectos Aleatorios

Wilmer Fermín¹, Haidee Castañeda²

¹Universidad de Oriente. Departamento de Estadística. Isla de Margarita-Venezuela.

wilmer_fermin@hotmail.com

²Universidad de Oriente, Núcleo de Anzoategui.

Resumen

La historia del análisis de regresión, el análisis de varianza (ANOVA) y el análisis de covarianza (ANCOVA) han escrito sus desarrollos separados; sin embargo, la apreciación y utilización de una notación matricial compartida permite englobarlos en el Modelo lineal general (GLM) con claras ventajas conceptuales y prácticas. El ANCOVA fue crucial en este tipo de generalización, puesto que éste combina características del análisis de regresión y el análisis de varianza para establecer posibles diferencias en la respuesta media entre grupos o tratamientos, fijados por el investigador, de uno o más factores, pero removiendo el posible efecto en la respuesta causado por la relación de ésta con una o más covariables (concomitantes). El ANCOVA de efectos aleatorios conocido como RANCOVA, se refiere a la situación en la que se tiene una población de niveles o tratamientos y el investigador tendrá que elegir aleatoriamente algunos para experimentar u observar la respuesta y la(s) concomitante(s); las conclusiones se infieren a toda la población de tratamientos. Por otra parte, el análisis multinivel es una metodología para analizar datos jerárquicos o multinivel con complejos patrones de variabilidad; la atención en el análisis está en las fuentes de variación anidadas. La herramienta fundamental del análisis multinivel es el modelo lineal jerárquico o multinivel, el cual representa una extensión del modelo de regresión múltiple que incluye dos o más términos de error: uno (o más) para cada nivel de la jerarquía y variables de cada nivel, cada nivel de la jerarquía está representado por un submodelo. En la literatura se puede notar el RANCOVA y el Análisis multinivel se presentan como si fuesen totalmente diferentes. No obstante, existe un enlace entre ambos métodos. En esta investigación el objetivo fue comparar el RANCOVA y el Análisis Multinivel. La metodología para hacer la comparación se basó en establecer las similitudes y diferencias entre ambos métodos estadísticos considerando sus fundamentos, supuestos, modelos y estimación de parámetros. El hallazgo más importantes es que el Análisis Multinivel es una generalización del Análisis de Covarianza de efectos aleatorios; se demostró que el modelo multinivel bajo una estructura jerárquica de dos niveles, el cual admite variables explicativas en ambos niveles y coeficientes aleatorios tanto en el nivel uno como en el nivel dos puede particularizarse al modelo del RANCOVA e incluso al ANCOVA. No obstante, la particularización debe hacerse en función de las consideraciones a los grupos o tratamientos y al interés de en los contrastes estadística.

Palabras clave: Ancova, rancova, análisis multinivel, comparación.

1. Introducción

El ANCOVA es una técnica estadística que combina la regresión y análisis de varianza para detectar posibles diferencias en la respuesta media entre grupos o tratamientos una vez que se corrige el efecto de variables concomitantes en la respuesta. Al igual que el análisis de varianza el ANCOVA puede ser de efectos fijos y de efectos aleatorios. A este último se le llama RANCOVA. Una de las grandes limitaciones del ANCOVA (RANCOVA) es que el efecto de las variables independientes sobre la respuesta es revelado (capturado) más precisamente cuando la influencia sobre las variables independientes representado por las concomitantes es igual a través de las condiciones experimentales

El análisis multinivel es una metodología para el análisis de datos jerárquicos o multinivel con complejos patrones de variación, con atención en las fuentes de variabilidad anidadas. Los datos multinivel se caracterizan por que unidades individuales (nivel uno) están anidadas en unidades superiores (nivel dos); por ejemplo, alumnos en escuelas, empleados en empresas, datos longitudinales, entre otros. La metodología considera la formulación de una función matemática para modelizar los datos tomando en cuenta esa estructura jerárquica y hacer posible la incorporación de variables de todos los niveles de la jerarquía. (Fermín, et. al., 2007; Goldstein, 1995; Snijders y Bosker, 1999). Los coeficientes de regresión varían de nivel a nivel, debido a que el efecto de las variables puede variar en las distintas unidades del nivel dos (grupos).

El objetivo de esta investigación fue comparar teóricamente el ANCOVA y el análisis multinivel. En cuanto a la metodología para hacer la comparación se buscó discernir respecto a las semejanzas, diferencias y equivalencias entre los modelos, los fundamentos y supuestos de cada método.

2. Modelo del Análisis de Covarianza de Efectos Aleatorios

Se considerará el modelo de una vía (una única variable independiente o factor) con “p” variables concomitantes (covariables). Explícitamente, se quiere estudiar el efecto de un factor, con un gran número de tratamientos o niveles, sobre una respuesta “Y” y p variables concomitantes relacionadas linealmente con Y pero el investigador no logra controlarlas; además, considérese que la influencia sobre el factor representado por las concomitantes es igual a través de las condiciones experimentales. Considérese el caso que el investigador elige aleatoriamente “m” tratamientos del factor para observar la respuesta y las concomitantes. Esto es motivo de análisis usando un RANCOVA que permite probar la hipótesis de igualdad de las medias de los tratamientos del factor una vez que se separa el efecto de las concomitantes sobre la respuesta. El modelo viene dado por:

$$Y_{ij} = \mu + \lambda_j + \sum_{s=1}^p \beta_s (X_{sij} - \bar{X}_s) + e_{ij} \quad \begin{cases} i = 1, \dots, n_j \\ j = 1, \dots, m \end{cases} \quad (1)$$

Los parámetros μ y β_s representa la media global, el efecto del j-ésimo tratamiento y el coeficiente de regresión de X_s . λ_j 's son variables aleatorias $IIDN(0, \sigma_\lambda^2)$. Se rechaza la

igualdad de medias de los tratamientos si la varianza σ_λ^2 resulta ser significativamente mayor que cero. Por otra parte se espera que $\beta_s \neq 0$ para algún $j = 1, \dots, p$.

Una de las grandes limitaciones del ANCOVA (RANCOVA) para efectos de aplicaciones prácticas es la suposición de que las covariables son independientes del factor (factores en el caso de considerar varios); es decir que las condiciones experimentales no afectan la distribución de las puntuaciones de las variables concomitantes. Rutherford (2001) ha sugerido una variedad de alternativas en el caso del infringimiento del supuesto de independencia entre las concomitantes y los factores.

3. Modelo Multinivel

A continuación se presenta una situación ideal para generalizar el modelo jerárquico lineal de dos niveles. Supóngase una estructura jerárquica de dos niveles con m grupos o unidades del nivel dos y n_j unidades del nivel uno dentro del j -ésimo grupo; el interés está en caracterizar una respuesta Y del nivel uno, en función de x_1, x_2, \dots, x_p (p variables explicativas del nivel uno) y z_1, z_2, \dots, z_p (q variables del nivel dos). Suponga, además, que de las p variables en el nivel uno: p_1 admiten coeficientes aleatorios en el nivel 1 y en el nivel 2 (x_1, x_2, \dots, x_{p_1}), p_2 admiten coeficientes aleatorios en el nivel 2 ($x_{p_1+1}, x_{p_1+2}, \dots, x_{p_2}$) y p_3 admiten coeficientes fijos ($x_{p_2+1}, x_{p_2+2}, \dots, x_{p_3}$), sabiendo que: $p = p_1 + p_2 + p_3$, $p_1 < p$, $p_2 < p$ y $p_3 < p$. Con estas consideraciones, el modelo jerárquico lineal de dos niveles se puede escribir como sigue:

$$y_{ij} = \sum_{h=0}^{p_2} \sum_{s=0}^q \lambda_{hs} Z_s x_{hij} + \sum_{h=p_2+1}^{p_3} \beta_h x_{hij} + \sum_{h=0}^{p_2} U_{hj} x_{hij} + \sum_{h=0}^{p_1} \varepsilon_{hij} x_{hij} \quad (2)$$

Donde $\sum_{h=0}^{p_2} \sum_{s=0}^q \lambda_{hs} Z_s x_{hij} + \sum_{h=p_2+1}^{p_3} \beta_h x_{hij}$ representa la curva promedio sobre todos los grupos, $\sum_{h=0}^{p_2} U_{hj} x_{hij}$ son las fluctuaciones o efectos de interacción de los grupos y las variables del nivel uno y $\sum_{h=0}^{p_1} \varepsilon_{hij} x_{hij}$ son las fluctuaciones o efectos individuales de las explicativas X 's en cada unidad dentro del grupo.

4. Análisis de Covarianza un Caso Particular del Análisis Multinivel

El ANCOVA tanto de efectos fijos como aleatorios son casos particulares del modelo jerárquico lineal general de dos niveles; veamos esto. Recordemos que el modelo multinivel general de dos niveles viene dado por (6). Además, se sabe que $p = p_1 + p_2 + p_3$; p_1 es el número de variables que admiten coeficientes aleatorios tanto en el nivel 1 como en el nivel 2; p_2 admiten coeficientes aleatorios en el nivel 2 y p_3 admiten coeficientes fijos.

Particularizando la situación al caso donde $p = p_3 = 1$ y $q = 0$; es decir, existe sólo una variable explicativa al nivel 1 y admite coeficiente fijo. Es claro que $q = 0$, $p_1 = p_2 = 0$; al sustituir y operar en la ecuación (6) se tiene

$$y_{ij} = \lambda_{00}Z_0x_{0ij} + \beta_1x_{1ij} + U_{0j}x_{0ij} + \varepsilon_{0ij}x_{0ij} \quad (3)$$

Pero, debido a que solo se está considerando una misma variable explicativa medida en los distintos factores y no existen características de los factores, es decir, $Z_0 = 1$ y $x_{0ij} = 1$, se obtiene el modelo

$$y_{ij} = \lambda_{00} + \beta_1x_{1ij} + U_{0j} + \varepsilon_{0ij} \quad \begin{cases} i = 1, 2, \dots, n_j \\ j = 1, 2, \dots, m \end{cases} \quad (4)$$

Esta ecuación es justamente el modelo de RANCOVA de efectos aleatorios, descrito en la ecuación (1) pero con una sola variable concomitante. En este caso se tiene $\lambda_{00} = \mu$ y $U_{0j} = \lambda_j$.

Como se puede ver, del modelo jerárquico lineal general se ha llegado al ANCOVA de efectos aleatorios con unas suposiciones apropiadas y que pueden darse en la realidad. Bajo las suposiciones establecidas, se puede presumir que existe un factor cuyos niveles o tratamientos son justamente los grupos o unidades del nivel 2, que estos son suficientemente numerosos que forman una población finita o infinita, que el investigador puede elegir aleatoriamente los tratamientos, que la covariable, X , está relacionada linealmente con la respuesta y que no está afectada por los tratamientos.

Por otra parte, si el número de grupos es pequeño (N pequeño) que pueda hacer irrelevante una selección aleatoria de éstos, se puede restringir a que U_{0j} sea un parámetro fijo, esto puede ser abordado con el modelo ANCOVA de efectos fijos. Snijders y Bosker (1999), plantean varias ventajas del análisis multinivel en relación al ANCOVA de efectos fijos. Entre estas se menciona el caso cuando el tamaño de los grupos son pequeños; en este caso el modelo de coeficiente aleatorio tiene una importante ventaja sobre el análisis de covarianza, siempre que las suposiciones acerca del coeficiente aleatorio sean razonable; es decir, el efecto de grupo U_{0j} , en el modelo de coeficiente aleatorio, se considera independiente e idénticamente distribuidos, lo que ayuda a contrarrestar la escasez de datos que implica tamaños de grupos relativamente pequeños, ya que todos los efectos de grupos se supone provienen de la misma población y los datos de cada grupo tienen relación con respecto a otros grupos; mientras que en el análisis de covarianza, cada U_{0j} es estimado como un parámetro separado, donde los datos no contienen mucha información acerca de los valores de U_{0j} .

5. Conclusiones

- (1) Análisis Multinivel es una generalización del Análisis de Covarianza de efectos aleatorios, cuya principal herramienta es el modelo Multinivel de dos niveles, el cual admite variables explicativas en ambos niveles y coeficientes aleatorios tanto en el nivel uno como en el nivel dos.
- (2) el análisis multinivel posee mayor alcance que el ANCOVA, ya sea con efectos fijos o efectos aleatorios, puesto que el modelo multinivel caracteriza la respuesta considerando que la variable concomitante puede estar relacionada con los factores y esta es la mayor limitación del ANCOVA.

REFERENCIAS

- Fermín, W., Galindo, P. y Martín, J. (2007). Estrategia multinivel para detectar comportamientos de abandonos en el análisis de datos longitudinales. *Saber*, Vol. 19 N° 1, 65-73.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 1, 43-56.
- Goldstein, H. 1995. *Multilevel Statistical Models*. Second Edition. London: Edward Arnold.
- Longford, N. (1993). *Random Coefficient Models*. Editorial Biddles Short Run Books Kings Lynn. Gran Bretaña.
- Rutherford A. (2001). *Introducing ANOVA and ANCOVA: A GLM Approach*. First edition. London. SAGE Publicación.
- Snijders, T. and Bosker, R. (1999). *An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Representación del BLUP en el Modelo Lineal General Mixto, Caso Balanceado

Fernando Velasco Luna¹, Mario Miguel Ojeda Ramirez²

¹Facultad de Estadística e Informática. UV. fvelasco@uv.mx

²Facultad de Estadística e Informática. UV. mojeda@uv.mx

Resumen

El modelo lineal general mixto engloba una clase amplia de modelos que permiten tratar situaciones con datos de estructura jerárquica. Es de vital importancia para realizar procesos de inferencia usando estos modelos contar con resultados de estimación y prueba de hipótesis para los efectos fijos y aleatorios. Aunque en la literatura se conocen suficientes resultados acerca de la teoría del álgebra lineal relacionada con la teoría de estimación y de prueba de hipótesis en el modelo lineal general (MLG), hace falta material que muestre la relación existente del álgebra lineal y la teoría de la predicción en el modelo lineal general mixto (MLGM) $Y = X\beta + Zu + e$. En este trabajo se considera la representación del Mejor Predictor Lineal Insesgado (*BLUP*) de un efecto mixto en términos del operador el proyector orthogonal $P_Z = Z(Z^t Z)^{-1} Z^t$ sobre el espacio $S(Z)$ y del operador proyector oblicuo $P_{X:XY^{-1}} = X(X^t V^{-1} X)^{-1} X^t V^{-1}$, sobre el espacio $S(X)$. Se presenta la condición que debe cumplir la matriz de diseño Z bajo la cual el *BLUP* de un efecto mixto se puede expresar en términos de los operadores mencionados. Se consideran casos particulares del modelo lineal general mixto con dos componentes de la varianza, con lo cual estos resultados se pueden aplicar a modelos particulares.

Palabras Clave: Componentes de la varianza, efecto mixto, modelación multinivel, modelo lineal jerárquico, operador proyector.

1. Introducción

Los modelos lineales jerárquicos forman una clase general que permite abordar situaciones con datos que presentan una estructura jerárquica, en diversas áreas de aplicación como: investigación educativa (efectividad de escuela, logro escolar), biología (curvas de crecimiento, estudios genéticos), investigación social (análisis de encuestas, estudios de mercado), psicología (análisis de conducta), medicina (ajuste de datos de medidas repetidas y estudios de centros hospitalarios), entre otras. Recientes desarrollos en cómputo han permitido que se incremente el uso de estos modelos, que son también conocidos como modelos multinivel (De Leeuw, 2008), modelos de componentes de la varianza y covarianza (Searle *et al.*, 2006), o como modelos de efectos mixtos (West *et al.*, 2007). Un tratamiento y abundantes referencias acerca de estos modelos se pueden encontrar en (De Leeuw, 2008, Gelman and Hill, 2007; West *et al.*, 2007).

La teoría de espacios vectoriales de dimensión finita proporciona un marco para trabajar conceptos de inferencia en el MLG $Y = X\beta + e$, $E(e) = 0$, $Cov(e) = \sigma^2 I$ donde $Y \in R^n$, X es una matriz de constantes de orden $n \times p$, $\beta \in R^p$ es un vector de parámetros desconocidos, y $e \in R^n$ es un vector de errores aleatorios no observables. Conceptos como subespacio columna, operador proyector ortogonal u oblicuo, matriz inversa generalizada, entre otros, juegan un papel de suma importancia en el estudio del MLG. Los proyectores asociados con un estimador en particular en el MLG juegan un papel importante en la caracterización de las propiedades estadísticas del estimador. En el criterio de mínimos cuadrados para definir el mejor ajuste, el estimador de mínimos cuadrados ordinarios del vector de coeficientes en el MLG se puede expresar en términos del operador proyector ortogonal $P_x = X(X^t X)^{-1} X^t$ sobre el espacio $S(X)$; así mismo el estimador de mínimos cuadrados generalizados se puede expresar en términos del operador proyector oblicuo $P_{X: XV^{-1}} = X(X^t V^{-1} X)^{-1} X^t V^{-1}$. Para una revisión de la aplicación del operador proyector en modelos lineales ver Christensen (2002). Kala y Pordzik, (2006) presentan propiedades de operadores lineales y su relación con el Mejor Estimador Lineal Insesgado (*BLUE*). Peng *et al.* (2005) presentan algunas propiedades del operador oblicuo.

Considérese el modelo $Y = X\beta + e$, $E(e) = 0$, $Cov(e) = \sigma^2 I$ donde $Y \in R^n$, X es una matriz de constantes de orden $n \times p$, $\beta \in R^p$ es un vector de parámetros desconocidos, y $e \in R^n$ es un vector de errores aleatorios no observables. Es conocido que $\hat{\beta}$ es un estimador mínimos cuadrados de β si, y sólo si, $X\hat{\beta} = P_x Y$. Si se considera el modelo $Y = X\beta + e$, $E(e) = 0$, $Cov(e) = \sigma^2 I$, donde V es una matriz definida positiva, entonces $\hat{\beta}$ es un estimador mínimos cuadrados de β si, y sólo si, $X\hat{\beta} = P_{X: XV^{-1}} Y$; en este caso $\hat{\beta}$ se denomina el estimador mínimos cuadrados generalizado de β (ver Christensen (2002) para demostraciones).

2. Mejor Predictor Lineal Insesgado y Estimación de Efectos Mixtos

Considérese el MLGM

$$Y = X\beta + Zu + e \quad (1)$$

donde $Y \in R^n$, X y Z son matrices de constantes de orden $n \times p$ y $n \times q$ respectivamente, $\beta \in R^p$ es un vector de parámetros fijos desconocidos, $u \in R^q$ es un vector de efectos aleatorios, $u \sim N(0, \sigma_u^2 I_m)$, y $e \in R^n$ es un vector de errores aleatorios no observables, $e \sim N(0, \sigma_e^2 I_n)$. La matriz de varianzas y covarianzas de Y está dada por:

$$V = Cov(Y) = \sigma_u^2 ZZ^t + \sigma_e^2 I. \quad (2)$$

En una serie de trabajos Henderson desarrolla el método del mejor predictor lineal insesgado para modelos mixtos. Concretamente Henderson *et al.* (1959) desarrollaron un conjunto de ecuaciones que simultáneamente proporcionan el *BLUE* de $X\beta$ y el *BLUP* de u . Estas ecuaciones son derivadas por la maximización de la densidad conjunta de Y y u , la cual está dada para, $Var(e) = R$ y $Var(u) = G$ por:

$$f(Y, u) = \frac{\exp\left\{-\frac{1}{2}[(Y - X\beta - Zu)^t R^{-1}(Y - X\beta - Zu) + u^t G^{-1}u]\right\}}{(2\pi)^{\frac{(N+q)}{2}} |R|^{\frac{1}{2}} |G|^{\frac{1}{2}}}. \quad (3)$$

Igualando a cero las derivadas parciales de (3) con respecto a los elementos de β y de u , se obtienen las ecuaciones del MLGM

$$\begin{bmatrix} X' R^{-1} X & X' R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X' R^{-1} Y \\ Z' R^{-1} Y \end{bmatrix}. \quad (4)$$

Las estimaciones pueden ser escritas como:

$$\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} Y; \quad \hat{u} = G Z' V^{-1} (Y - X \hat{\beta}). \quad (5)$$

En este contexto es necesaria la estimación de funciones de la forma $k^t \beta + m^t u$, para vectores específicos de constantes κ y m . Henderson (1975) obtiene el *BLUP* de $k^t \beta + m^t u$ que está dado por:

$$BLUP(k^t \beta + m^t u) = k^t \hat{\beta} + m^t \hat{u}. \quad (6)$$

A continuación se presenta la condición que debe cumplir la matriz de diseño Z bajo la cual el *BLUP* del efecto mixto $X\beta + Zu$ se puede expresar en términos de los operadores $P_{X:VX^{-1}}$, y P_Z , sobre los subespacios $S(X)$ y $S(Z)$ respectivamente.

Teorema 1. Bajo el modelo (1), si existe una constante d tal que $ZZ^t = dP_Z$, entonces la inversa V^{-1} de V se puede expresar en términos del operador proyector ortogonal P_Z y de su complemento Q_Z por medio de $V^{-1} = \frac{P_Z}{(d\sigma_u^2 + \sigma_e^2)} + \frac{Q_Z}{\sigma_e^2}$.

Demostración: Se tiene que bajo el modelo (1), si existe una constante d tal que $ZZ^t = dP_Z$, entonces

$$\begin{aligned} V &= \sigma_u^2 ZZ^t + \sigma_e^2 I = d\sigma_u^2 P_Z + \sigma_e^2 (P_Z + Q_Z) \\ &= d\sigma_u^2 P_Z + \sigma_e^2 P_Z + \sigma_e^2 Q_Z = (d\sigma_u^2 + \sigma_e^2) P_Z + \sigma_e^2 Q_Z, \end{aligned}$$

de lo cual la inversa V^{-1} se puede expresar por medio de $V^{-1} = \frac{P_Z}{(d\sigma_u^2 + \sigma_e^2)} + \frac{Q_Z}{\sigma_e^2}$.

Teorema 2. Bajo el modelo (1), si existe una constante d tal que $ZZ^t = dP_Z$, entonces el *BLUP* del efecto mixto $X\beta + Zu$ se puede expresar en términos de los operadores $P_{X:VX^{-1}}$ y P_Z , por medio de $P_{X:VX^{-1}} Y + c P_Z Y - c P_Z P_{X:VX^{-1}} Y$, donde $c = \frac{d\sigma_u^2}{(d\sigma_u^2 + \sigma_e^2)}$.

Demostración: Por la teoría de Henderson el *BLUP* de $X\beta + Zu$ está dado por medio de $X\hat{\beta} + Z\hat{u}$ donde $\hat{\beta}$ es el estimador de mínimos cuadrados generalizados de β y \hat{u} es el *BLUP* de u . Se sabe que el estimador mínimos cuadrados generalizado se relaciona con el operador proyector $P_{X:VX^{-1}}$ por medio de $X\hat{\beta} = P_{X:VX^{-1}} Y$, y que $\hat{u} = GZ^t V^{-1} (Y - X\hat{\beta})$. Así el *BLUP* de $X\beta + Zu$ se puede expresar como:

$$\begin{aligned}
X\hat{\beta} + Z\hat{u} &= P_{X:VX^{-1}}Y + ZGZ'V^{-1}\left(Y - X\hat{\beta}\right) \\
&= P_{X:VX^{-1}}Y + \sigma_u^2ZZ'\left(\frac{P_Z}{(d\sigma_u^2 + \sigma_e^2)} + \frac{Q_Z}{\sigma_e^2}\right)\left(Y - X\hat{\beta}\right) \\
&= P_{X:VX^{-1}}Y + d\sigma_u^2P_Z\left(\frac{P_Z}{(d\sigma_u^2 + \sigma_e^2)} + \frac{Q_Z}{\sigma_e^2}\right)\left(Y - X\hat{\beta}\right) \\
&= P_{X:VX^{-1}}Y + \frac{d\sigma_u^2P_Z}{(d\sigma_u^2 + \sigma_e^2)}\left(Y - P_{X:VX^{-1}}Y\right) \\
&= P_{X:VX^{-1}}Y + cP_ZY - cP_ZP_{X:VX^{-1}}Y,
\end{aligned}$$

donde $c = \frac{d\sigma_u^2}{(d\sigma_u^2 + \sigma_e^2)}$.

3. Aplicaciones a Algunos Modelos Particulares

Casos particulares del MLGM que satisfacen la condición $ZZ' = dP_Z$ son: (a) *Modelo solo intercepto sin variables explicatorias*. Este es el caso más simple de un modelo lineal jerárquico; el modelo para la i -ésima unidad de nivel 1 en la j -ésima unidad de nivel 2 está dado por:

$$y_{ij} = \mu + u_j + e_{ij}, \quad i = 1, \dots, d, \quad j = 1, \dots, k, \quad (7)$$

donde μ es un parámetro fijo u_j ; es el efecto aleatorio; u_j y e_{ij} son independientes, con $u \sim N(0, \sigma_u^2)$ y $e \sim N(0, \sigma_e^2)$. El modelo para la j -ésima unidad de nivel 2 es

$$y_j = 1_d\mu + 1_du_j + e_j, \quad j = 1, \dots, k, \quad (8)$$

tomando $\beta = \mu$, $X = 1_k \otimes 1_d$, $Z = I_k \otimes 1_d$ y $u = (u_1, \dots, u_k)'$ el modelo (8) es de la forma $y = X\beta + Zu + e$. En este caso se tiene que:

$$\begin{aligned}
ZZ' &= (I_k \otimes 1_d)(I_k \otimes 1_d)' \\
&= (I_k \otimes 1_d)(I_k' \otimes 1_d') \\
&= (I_k I_k' \otimes 1_d 1_d') \\
&= (I_k \otimes 1_d 1_d');
\end{aligned}$$

así:

$$\begin{aligned}
P_Z &= Z(Z'Z)^{-1}Z' = (I_k \otimes 1_d) \left[(I_k \otimes 1_d)' (I_k \otimes 1_d) \right]^{-1} (I_k \otimes 1_d)' \\
&= (I_k \otimes 1_d) \left[(I_k' \otimes 1_d') (I_k \otimes 1_d) \right]^{-1} (I_k \otimes 1_d)' = (I_k \otimes 1_d) \left[(I_k' I_k \otimes 1_d' 1_d) \right]^{-1} (I_k \otimes 1_d)' \\
&= (I_k \otimes 1_d) \left[(I_k \otimes 1/d) \right] (I_k \otimes 1_d)' = \left[(I_k \otimes 1_d / d) \right] (I_k \otimes 1_d)' \\
&= I_k \otimes \frac{1_d 1_d'}{d} = \frac{1}{d} (I_k \otimes 1_d 1_d') = \frac{1}{d} ZZ',
\end{aligned}$$

cumpléndose la condición $dP_Z = ZZ'$.

(b) *Modelo sólo intercepto con variables explicatorias.* Considérese el modelo:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_j + e_{ij}, \quad i = 1, \dots, d, \quad j = 1, \dots, k, \quad (9)$$

donde β_0 , β_1 y β_2 son parámetros fijos; u_j es el efecto aleatorio; u_j y e_{ij} son independientes, con $u_j \sim N(0, \sigma_u^2)$ y $e_{ij} \sim N(0, \sigma_e^2)$. El modelo para la j -ésima unidad de nivel 2 tiene la forma

$$y_j = 1_d \beta_0 + X_{1j} \beta_1 + X_{2j} \beta_2 + 1_d u_j + e_j, \quad j = 1, \dots, k, \quad (10)$$

tomando $\beta = (\beta_0, \beta_1, \beta_2)'$, $X_j = (1_d : X_{1j} : X_{2j})$, $Z_j = 1_d$, el modelo (10) toma la forma $y_j = X_j \beta + Z_j u_j + e_j$, $j = 1, \dots, k$. Definiendo a X_1 y a X_2 como las matrices de las variables explicatorias, y $X = (1_k \otimes 1_d : X_1 : X_2)$, $Z = I_k \otimes 1_d$ y $u = (u_1, \dots, u_k)'$ el modelo (10) es de la forma $y = X\beta + Zu + e$, y se cumple la condición $ZZ' = dP_Z$.

(c) *Modelo de dos vías.* Este modelo es de la forma

$$y_{ij} = \mu + \alpha_j + \beta_i + e_{ij}, \quad i = 1, \dots, d, \quad j = 1, \dots, k, \quad (11)$$

donde μ y β_i son parámetros fijos; α_j es el efecto aleatorio; α_j y e_{ij} son independientes, con $\alpha_j \sim N(0, \sigma_\alpha^2)$ y $e_{ij} \sim N(0, \sigma_e^2)$. Definiendo $\beta = (\mu, \beta_1, \dots, \beta_d)'$, $X_j = (1_d : I_d)$, $Z_j = 1_d$ y $u_j = \alpha_j$ el modelo para la j -ésima unidad de nivel 2 está dado por medio de $y_j = X_j \beta + Z_j u_j + e_j$, $j = 1, \dots, k$. Tomando $\beta = (\mu, \beta_1, \dots, \beta_d)'$, $u = (\alpha_1, \dots, \alpha_k)$, $X = (1_k \otimes 1_d : 1_k \otimes I_d)$ y $Z = I_k \otimes 1_d$ el modelo (11) está dado por medio de $y = X\beta + Zu + e$. En este caso también se cumple la condición $ZZ' = dP_Z$.

4. Aplicación a datos reales

La representación del *BLUP* obtenida en términos de los operadores se aplico a datos relacionados con las calificaciones de estudiantes, se obtuvo el *BLUP* del efecto mixto en términos del operador proyector, y se comparo con el *BLUP* obtenido en base al resultado de Henderson. Los datos GPA son un conjunto de datos longitudinales, donde 100 estudiantes de una escuela han tenido seguimiento durante los 6 semestres de su escolaridad. El modelo usado fue el modelo solo intercepto sin variables explicatorias:

$$y_{ij} = \mu + u_j + e_{ij}, \text{ con } u_j \sim N(0, \sigma_u^2) \text{ y } e_{ij} \sim N(0, \sigma_e^2).$$

Se realizó la predicción del efecto mixto $X\beta + Zu$ obtenida en base a la teoría de Henderson del BLUP $X\hat{\beta} + Z\hat{u}$ y a la basada en proyectores $P_{X:VX^{-1}}Y + cP_ZY - cP_ZP_{X:VX^{-1}}Y$, encontrándose que no existe diferencia entre las predicciones obtenidas.

5. Conclusiones

Se hace la representación del mejor predictor lineal insesgado del efecto mixto $X\beta + Zu$, basado en el MLGM, en términos de los operadores $P_{X:VX^{-1}}$ y P_Z , sobre los subespacios $S(X)$ y $S(Z)$, respectivamente. La predicción obtenida se compara con la predicción obtenida basada en la teoría de Henderson, haciendo uso del modelo solo intercepto sin variables explicatorias, encontrándose que no existe diferencia entre las predicciones obtenidas. Esta representación permitirá poder estudiar desde el punto de vista del álgebra lineal la teoría de predicción en el MLGM y las propiedades distribucionales de los predictores.

REFERENCIAS

- Christensen, R. (2002). *Plane Answers to Complex Questions. The Theory of Linear Models*. 3rd Ed. New York: Springer.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423-447.
- Henderson, C.R., Kempthorne, O., Searle, S.R. and von Krosigk, C.N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics*. 15, 192-218.
- Kala, R. and Pordzik. (2006). Two operators and the BLUE. *Linear Algebra and its Applications* 417, 134-139.
- De Leeuw, J. and Meijer, E. (2008). *Handbook of Multilevel Analysis*. New York: Springer.
- Peng, C. and Zhang, X. (2005). On recursive oblique projectors. *IEEE Signal Processing Letters* 12, 433-436.
- Searle, S.R., Casella, G. and McCulloch, C.E. (2006). *Variance Components*. New York: John Wiley.
- West B. T., Welch K. B., and Galecki, A. T. (2007). *Linear Mixed Models. A Practical Guide Using Statistical Software*. Boca Raton: Chapman & Hall/CRC.

Evaluación de Insectidas para el Control de Garrapatas (*Boophilus microplus*) que Afectan al Ganado Bovino de Doble Propósito usando Modelos Lineales Generalizados

Rafael Borges¹, Andrea González², Miguel Pérez³, Mandius Carvajalino³, Daira Velandia³, Dilia Tapias²

¹ Universidad de los Andes. Escuela de Estadística. Mérida-Venezuela. borgesr@ula.ve

² Universidad Experimental del Sur del Lago. Santa Barbara. Zulia-Venezuela

³ Universidad de los Andes. Instituto de Estadística Aplicada y Computación. Mérida-Venezuela

Resumen

En el presente trabajo se presentan los resultados de la evaluación de *in vitro* de cuatro tipos de químicos utilizados en el control de garrapatas: Alfacipermetrina Amitraz, Cipermetrina y Ethion, sobre las garrapatas (*Boophilus microplus*) que afectan a animales bovinos de doble propósito. Las garrapatas se han recolectado en dos unidades de producción por cada parroquia del municipio Colon, Estado Zulia, Venezuela. Luego de una selección de ejemplares comparables se someten a inmersión en distintas concentraciones a partir de las dosis terapéuticas (1X o concentración recomendada por el fabricante) se calculan las concentraciones de las soluciones del garrapaticida concentraciones 0.25X, 0.5X, 1X, 2X y 4X de cada principio activo. Posteriormente se espera al momento de la eclosión de los huevos y se determina el número de larvas vivas y muertas. Los datos fueron analizados mediante técnicas de Modelos Lineales Generalizados, considerando primero modelos binarios de dosis-respuesta basados en la distribución binomial para modelar la proporción de larvas muertas para cada concentración, planteándose un modelo para cada insecticida. En esta fase se consideraron las tres funciones de enlace utilizadas: la logit, la probit y la log-log complementaria y se determina el mejor modelo en cada caso y se seleccionó el mejor modelo basado en los desvíos (deviances) y en criterios de información, usándose el criterio de Akaike (AIC), el χ^2 de Pearson, la devianza (deviance) y una adaptación del R^2 para el modelo binomial. Posteriormente, y luego de determinar la dosis optima para cada garrapaticida se plantea un modelo basado en tablas de contingencia para la dosis recomienda por el fabricante para cada producto, utilizando modelos basados en la distribución de Poisson para comparar los distintos insecticidas. Los datos fueron analizados utilizando el lenguaje R.

Palabras clave: Modelos lineales generalizados, modelos de respuesta binaria, modelos para tablas de contingencia, evaluación de insecticidas, garrapatas (*Boophilus microplus*).

Abstract

In this paper, we present the results of an evaluation of several chemical products used as insecticide: Alfacipermetrina Amitraz, Cipermetrina, and Ethion over the control of *Boophilus microplus* that affect the double purpose bovines. The insects (*Boophilus microplus*) were collected in two production units in each county of the municipality of Colon, in the Zulia State in Venezuela. After a selection of comparable insects, they were immersed in concentrations obtained from the recommended dose (1X) and concentration of the insecticides of 0.25X, 0.5X, 1X, 2X and, 4X of each active principles was used. Then, the experiment was followed until the birth of the eggs and the number of death and

alive larva was counted in each case. The data was analyzed using techniques related with Generalized Linear Models, considering first, binary models of dose-response based on the binomial distribution to model the proportion of death larva for each concentration, using models for each insecticide and, the “best” model was selected using the deviances and information criteria, considering the Akaike criteria (AIC) the Pearson χ^2 , the deviance and a adaptation of the R^2 for binomial model. After determining the best performance concentration of each insecticide, it was performed an analysis of contingency tables using a Poisson distribution to compare the effect of the insecticides, using in this case the recommended dose for each product. The data was analyzed using the language R.

Keywords: Generalized linear models, binary response (dose-response) models, models for contingency tables, evaluation of insecticides, *Boophilus microplus*.

1. Introducción

Los modelos lineales generalizados son una generalización de los modelos lineales y fueron unificados en una clase de modelos donde la distribución de la variable respuesta, y por ende de los errores puede ser un miembro de la familia exponencial y fueron propuestos originalmente en el trabajo de Nelder y Wedderburn (Nelder y Wedderburn ,1972). En esta clase de modelos se reúnen una serie de modelos presentados en la literatura anteriormente y que aparentemente no tenían conexión entre ellos.

Uno de las aplicaciones clásicas de los modelos lineales generalizados son los ensayos de dosis-respuesta, cuyos objetivos principales son modelar la probabilidad de éxito de un fenómeno y determinar la dosis efectiva de un producto. Para detalles adicionales pueden consultarse libros como el clásico de McCullagh y Nelder (McCullagh y Nelder, 1989) o el de Cordeiro y Demetrio (Cordeiro y Demetrio, 2007).

En este trabajo se presenta la evaluación de cuatro productos: Alfacipermetrina Amitraz, Cipermetrina y Ethion sobre la proporción de muertes de larvas de garrapatas (*Boophilus microplus*) recolectadas de dos unidades de producción por cada parroquia del municipio Colon, Estado Zulia, Venezuela, proponiéndose la modelización de la probabilidad de muerte de las larvas mediante modelos basados en la distribución binomial.

Adicionalmente, se presenta un análisis preliminar de una tabla de contingencia, basado en la distribución de Poisson para comparar los distintos insecticidas, considerándose solo la dosis recomendada por el fabricante.

2. Materiales y Métodos

El material experimental se realizo a través de la recolección al azar de garrapatas (*Boophilus microplus*) en de dos unidades de producción por cada parroquia del municipio Colon, Estado Zulia, Venezuela.

Luego de la recolección de las garrapatas, se construyeron grupos comparables en cuanto al peso de las garrapatas y cada grupo se sometió a un proceso de inmersión a uno a distintos productos en la concentración o dosis recomendad por el fabricante (1X) y en dosis

equivalentes a la cuarta parte de la dosis recomendada (0.25X), la mitad de la dosis recomendada (0.5X), el doble de la dosis recomendada (2X) y cuatro veces la dosis recomendada (4X). Luego de realizar este proceso, se las garrapatas se depositaron en una incubadora durante 15 días, esperándose hasta la eclosión de los huevos y se contabilizó el número de larvas vivas y muertas. Este procedimiento se realizó de manera similar y con otras garrapatas para cada uno de los productos utilizados.

En este trabajo, el éxito fue considerado como la muerte de las larvas de garrapatas y la modelización se llevó a cabo en el contexto clásico de los modelos de dosis-respuesta, caso especial de los modelos lineales generalizados, utilizando el modelo logit (modl), el modelo probit (modp) y el modelo log-log complementario (modc).

La selección del mejor modelo para cada insecticida fue realizada utilizando los siguientes criterios: el criterio de Akaike (AIC), el χ^2 de Pearson, la devianza (deviance) y una adaptación del R^2 para el modelo binomial.

Adicionalmente, se efectuó un análisis de una tabla de contingencia de doble entrada considerándose por un lado el producto y por el otro la condición de la larva (viva o muerta), efectuándose posteriormente un análisis de la tabla de contingencia, a través de un análisis log-lineal basado en la distribución de Poisson. Este procedimiento se realizó únicamente para la dosis recomendada por el fabricante para cada producto.

Los análisis presentados en el presente trabajo se llevaron a cabo utilizando el lenguaje de procesamiento estadístico R (R Development Core Team, 2009), utilizando adaptaciones del código computacional propuesto por Faraway (Faraway, 2006).

3. Resultados

La figura 1 muestra los resultados del ajuste para cada uno de los modelos y cada uno de los insecticidas evaluados, en estas representaciones se puede observar que para los tres primeros productos, las estimaciones de las probabilidad de muerte de las larvas es similar para cada uno de los modelos, quizás con un poco de diferencia para dosis más elevadas en el caso de Amitraz y Cipermetrina. Para el caso de Ethion, se observa que el modelo log-log complementario tiene un comportamiento ligeramente diferente al de los otros dos modelos, sobre todo desde las dosis intermedias en adelante.

Otro aspecto que hay que destacar es que para el caso del producto Alfacipermetrina no se observa el comportamiento típico de un insecticida por ser el patrón de respuesta decreciente al aumentar la dosis, este aspecto llama la atención y sugiere la necesidad de efectuar otros estudios para determinar qué es lo que pudiera estar ocurriendo con este producto.

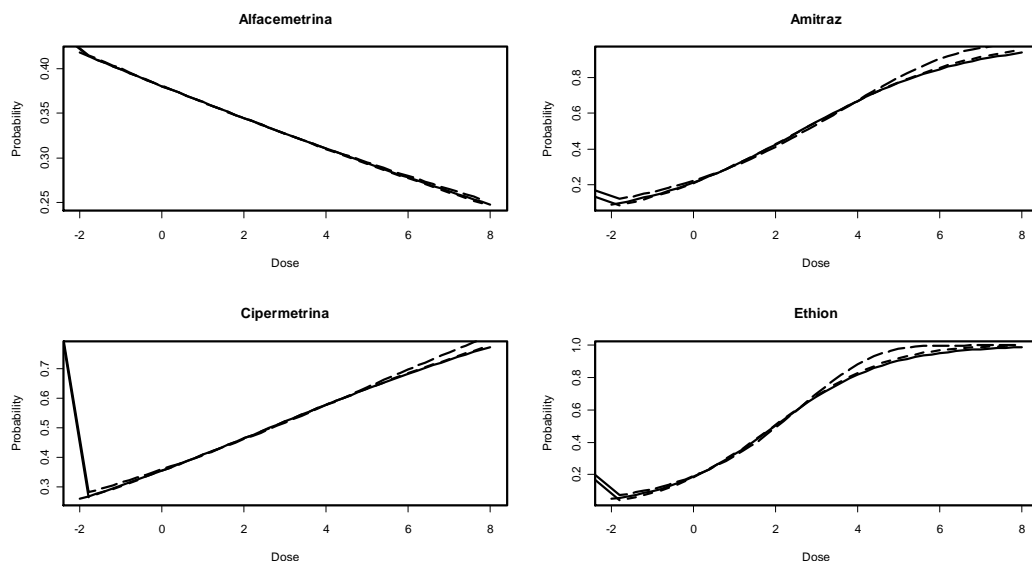


Figura 1. Valores ajustados para cada uno de los insecticidas en sus distintas dosis.

La tabla 1 muestra las medidas de calidad del modelo, basada en la bondad del ajuste de cada modelo para cada insecticida. En esta tabla se puede observar que el comportamiento de los modelos es similar para los tres primeros productos y que para el caso del Ethion el modelo log-log complementario presenta un mejor ajuste. Adicionalmente, se observa que el ajuste para el producto Alfacipermetrina es muy malo, lo que corrobora, lo observado en la figura 1.

Tabla 1
Criterio de Información de Akaike (AIC), χ^2 de Pearson, devianza (deviance) y R^2 adaptado para el modelo binomial cada uno de los insecticidas.

| Medida | Alfacetrina | | | Amitraz | | |
|----------|-------------|----------|----------|-----------|----------|----------|
| | modl | modp | modc | modl | modp | modc |
| AIC | 119 | 119 | 119.1 | 99.73 | 100 | 101.6 |
| χ^2 | 80.34303 | 80.32284 | 80.38507 | 63.21094 | 63.6198 | 64.29845 |
| Deviance | 84.48813 | 84.4713 | 84.52328 | 63.37424 | 63.67674 | 65.26133 |
| R^2 | 0.04861614 | 0.048810 | 0.048211 | 0.8351302 | 0.834333 | 0.830158 |

| Medida | Cipermetrina | | | Ethion | | |
|----------|--------------|----------|----------|-----------|----------|----------|
| | modl | modp | modc | modl | modp | modc |
| AIC | 57.98 | 57.89 | 59.14 | 142.1 | 140.4 | 114.1 |
| χ^2 | 20.30067 | 20.21173 | 21.45473 | 99.04742 | 98.52018 | 77.6498 |
| Deviance | 20.34793 | 20.25825 | 21.51185 | 109.5322 | 107.8058 | 81.54943 |
| R^2 | 0.6920535 | 0.693415 | 0.674376 | 0.7321054 | 0.736473 | 0.802311 |

En la primera parte de la figura 2 se puede observar que para la dosis recomendada, el insecticida con el que se obtiene una probabilidad de muerte más elevada es con el Cipermetrina, sin embargo, solo se obtienen probabilidades del orden de 0.4. En este gráfico, se observa que para dosis mayores, el Ethion tiene mejores efectos sobre la mortalidad de las larvas. En la segunda y tercera parte de la figura 2 se puede observar que para la dosis recomendada por el fabricante, la Cipermetrina es el producto con el que se

obtienen mejores resultados, y en la cuarta parte se puede observar que este producto está asociada con la muerte de las larvas para esta concentración.

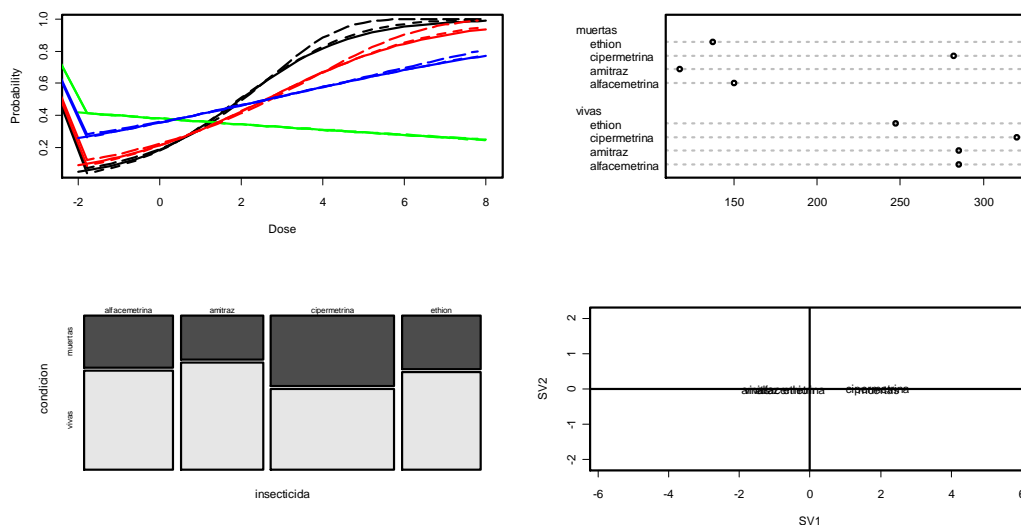


Figura 2. Valores ajustados para cada uno de los insecticidas en sus distintas dosis. Gráfico de puntos (Dotchart) para vivos y muertos por insecticida. Gráfico de mosaico (Mosaic Plot) para Insecticida y condición de la larva y Biplot del análisis de correspondencia entre insecticida y condición de la larva.

5. Referencias

6.

Cordeiro, G.M. y Demetrio, C.G.B. (2007). *Modelos Lineales Generalizados*. Universidade Federal de Santa Maria, RS.

Faraway, J.J. (2006). *Extending the Linear Models with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/ CRC, Boca Ratón, Florida.

McCullagh, P. y Nelder, J.A. (1989). *Generalized Linear Models*. Segunda edición. Londres: Chapman and Hall.

Nelder, J. A. and Wedderburn, R.W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, **135**, 370-384.

R Development Core Team (2009). R: A language and environment for statistical computing. Viena, Austria: R Foundation for Statistical Computing.

Uso de los Modelos Monofásico, Difásico y Gamma Incompleta para Caracterizar Curvas de Lactancia en Cabras Lecheras

Castañeda-Bustos Vielka Jeanethe¹, Torres-Hernández Glafiro^{1*}, Hernández-Mendo Omar¹, Mandeville Peter B.², Díaz-Gómez Marta Olivia², Ochoa-Cordero Manuel Antonio², Pérez-Elizalde Sergio¹, González-Camacho Juan Manuel¹

¹ Campus Montecillo-Colegio de Postgraduados. Montecillo, Edo. de México. México

² Universidad Autónoma de San Luis Potosí. San Luis Potosí. México

*glatohe@colpos.mx

Resumen

La producción de leche de cabra constituye una fuente muy importante de ingresos para los productores en las zonas áridas y semiáridas de México. Sin embargo, hay información limitada en México sobre caracterización de las curvas de producción de la leche, por lo que se realizó el presente estudio con el objetivo de caracterizar la curva de lactancia de cabras F₁ (Alpina x Nubia) del Altiplano Potosino. Se evaluaron los modelos monofásico, difásico y gamma incompleta, considerando como criterios de mejor ajuste: CME, pseudo R², AIC, BIC y -2 log likelihood. El modelo difásico fue el que mejor ajustó los datos al ser el que tuvo más consistencia en todos los criterios de ajuste, por lo que resultó más útil que los modelos monofásico y gamma incompleta para la caracterización de la curva de producción láctea.

Palabras clave: funciones matemáticas, curvas de lactancia, cabras, modelo multifásico, gamma incompleta.

1. Introducción

La representación gráfica de la producción de leche a través del tiempo, es una curva de lactancia (Sherchand *et al.*, 1995), su objetivo principal es predecir la producción diaria con el mínimo error, en presencia de diversos factores (Lombaard, 2006). El conocimiento de esta curva es una herramienta valiosa, ya que puede ser utilizada para diferentes aspectos de manejo (Gipson *et al.*, 1990), alimentación (Sherchand *et al.*, 1995) y genética (Weigel *et al.*, 1992). Esta curva ha sido modelada con diversas ecuaciones matemáticas (Tozer y Huffaker, 1999), pero la utilidad de cada una, dependerá de que tan bien describan o imiten el proceso biológico de la lactancia (Olori *et al.*, 1999), evaluándolas mediante ciertos criterios de ajuste como son: Cuadrado Medio del Error (CME), Pseudo - R² (Schabenberger, 2008), Criterio de información de Akaike (AIC) y Bayesiano (BIC); y -2 Log likelihood (Juneja *et al.*, 2006). Las funciones más utilizadas son: gamma incompleta, polinomio inverso, polinomio segmentado, y multifásica. Esta última, divide la curva en fases que proporcionan información útil de las características de la lactancia como: producción inicial y máxima; tiempo de la máxima producción y persistencia (Gipson y Grossman, 1990), ajusta mejor que la función gamma incompleta (Grossman y Koops, 1988) y permite desarrollar factores de ajuste para corregir efectos ambientales (De Boer *et al.*, 1989). A pesar de que la producción caprina constituye una fuente importante de ingresos en zonas áridas y semiáridas de México (Mellado *et al.*, 2004), no se ha caracterizado una curva para producción o componentes lácteos, por lo que se realizó el

presente estudio con el objetivo de determinar las curvas de lactancia y componentes lácteos de cabras F₁ (Alpina x Nubia) del altiplano potosino.

2. Materiales y Métodos

El estudio se realizó en la Unidad Caprina, Facultad de Agronomía, Universidad Autónoma de San Luis Potosí (UASLP), altitud de 1833 m, precipitación media anual: 335mm, clima seco estepario frío (García, 1973). Se utilizaron datos de la producción láctea de 18 cabras F₁ (Alpina x Nubia) de segundo parto mantenidas en estabulación y alimentadas durante el último tercio de gestación y el periodo de muestreo con 2 kg MS, aportada por 3.7 kg de alfalfa verde y 1 kg en base fresca de un concentrado con 16% PC y 2.4 Mcal; el concentrado se proporcionó en partes iguales en la mañana y en la tarde; el forraje se proporcionó únicamente en la mañana. Las cabras se ordeñaron una vez al día durante la mañana. La medición de la producción se efectuó una vez por semana, durante 14 semanas. Se obtuvieron las medias, desviación estándar, coeficiente de variación (CV), producción máxima y producción mínima. Las funciones evaluadas fueron: monofásica (Sherchand *et al.*, 1995), difásica (Gipson y Grossman, 1989) y gamma incompleta (Tozer y Huffaker, 1999).

$$\text{Función monofásica: } Y_t = a_i b_i [1 - \tanh^2(b_i(t - c_i))]]$$

$$\text{Función difásica: } Y_t = \sum_{i=1}^2 \{ a_i b_i [1 - \tanh^2(b_i(t - c_i))] \}$$

Donde: Y_t = producción de leche en un tiempo t ; a_i = promedio del total de producción por fase (kg); b_i = tasa de producción de a_i (por día) por cada fase i ; c_i = tiempo del pico de producción (días) por cada fase i ; t = tiempo (días); \tanh = función hiperbólica de la tangente.

$$\text{Función gamma incompleta: } Y_t = at^b e^{-ct}$$

Donde el parámetro a determina la producción inicial y el nivel máximo de producción, mientras que los parámetros b y c determinan la forma de la curva, y cómo cambia a través de la lactancia.

Los parámetros y criterios de ajuste se obtuvieron mediante PROC NLMIXED (SAS, 2002). Considerando: CME, Pseudo-R², AIC, BIC y -2 Log likelihood para la comparación de los modelos. Se seleccionó como mejor modelo aquel que fuera más consistente en éstos.

- AIC (Juneja *et al.*, 2006): $AIC = n \ln \left(\frac{SSE}{n} \right) + 2p$
- BIC (Juneja *et al.*, 2006): $BIC = n \ln \left(\frac{SSE}{n} \right) + 2(p + 2)q - 2q^2$

Donde n es el número de observaciones; SSE es la suma de cuadrados del error; p el número de parámetros en el modelo; $q = n\hat{\sigma}^2 / SSE$ y $\hat{\sigma}^2$ es el error de varianza estimado del modelo ajustado.

- Pseudo - R^2 (Schabenberger, 2008) definido por:

$$Pseudo - R^2 = 1 - SC(Residual) / SC(Total_{Corregido})$$

Las características y la curva de lactancia se determinaron con el modelo de mejor ajuste, calculados mediante a/b_i , c_i y $2/b_i$ para las funciones monofásica y difásica (Gipson y Grossman, 1990); y con $a(b/c)^b e^{-b}$, b/c y $-(b+1)\ln(c)$ para la función gamma incompleta (Rekik y Ben, 2004); para producción al pico, tiempo en que ocurre el pico, y “persistencia”; respectivamente.

3. Resultados y Discusión

La producción láctea presentó dos picos de producción, es decir, atípico. Las mayores producciones se observaron en la semana 2 (1133.89 ± 236.47 ml) y en la semana 12 (1152.22 ± 232.54 ml). Este comportamiento se atribuye a un periodo corto de problemas en la disponibilidad de alimento. El periodo más crítico ocurrió en la semana 8 (845.22 ± 172.86 ml). De acuerdo con los criterios de ajuste, coincidimos con Scott *et al.* (1996) en que el modelo que mejor simuló la lactancia fue el difásico (Figura 1); ya que al estar diseñado para modelar múltiples picos (Weigel *et al.*, 1992) funciona bien en caso de curvas atípicas como la aquí observada; mientras que la gamma incompleta no es útil en estos casos Rekaya *et al.* (2001) y Faro y Albuquerque (2002), al sobrestimar la producción (Tozer y Huffaker, 1999; Faro y Albuquerque, 1999) al inicio de la lactancia.

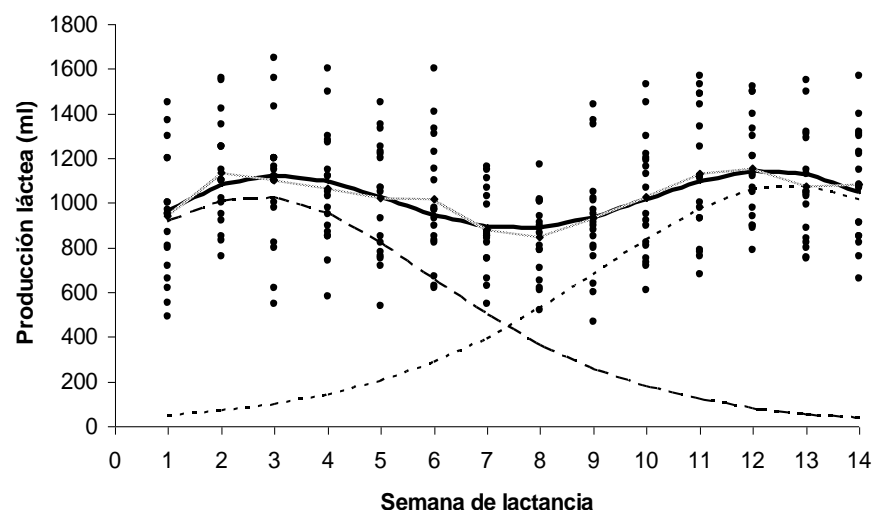


Figura 1. Producción láctea observada (•), promedio (.....) y predicha mediante la función difásica [general (—), fase 1 (---) y fase 2 (-·-)] para cabras F₁ (Alpina x Nubia).

El tiempo al pico de la lactancia (Cuadro 1) fue inferior a los encontrados por Gipson y Grossman (1989) (fase 1= día 37 y fase 2= día 113) y Montaldo *et al.* (1997) (día 56); pero similar al encontrado por Ruvuna *et al.* (1995) (día 28 de lactancia). Es necesario enfatizar que el genotipo aquí estudiado es una cruce Alpina x Nubia de segundo parto, y los valores pueden variar por el genotipo (Gonçalves *et al.*, 2002) y número de parto (Epaphras *et al.*, 2004; Miranda y Schnitkey, 1995), entre otros. La duración de cada fase tuvo una considerable diferencia con los resultados de Gipson y Grossman (1989) (raza Nubia= 274 y 952 d; raza Alpina= 179 y 606 d para las fases 1 y 2; respectivamente). Disparidad que puede deberse a que incluyeron lactancias de 305 días, lo cual no sucedió en el presente estudio.

Cuadro 1. Características de la lactancia de cabras F₁ (Alpina x Nubia).

| Modelo | Producción al pico* (ml) | Característica | |
|--------------------------|-----------------------------|----------------------------|---------------------------|
| | | Tiempo al pico** (días) | Persistencia*** (días) |
| <i>Producción láctea</i> | | | |
| Difásica ¹ | | | |
| Fase 1 | 1029.97 | 27 | 9.61 |
| Fase 2 | 1080.57 | 127 | 10.42 |

$$Y_t = \sum_{i=1}^2 \{a_i b_i [1 - \tanh^2(b_i(t - c_i))]\} \quad (* a_i b_i, **c, *** 2/b_i)$$

4. Conclusiones

El modelo difásico es útil en la descripción de lactancias atípicas como la aquí estudiada, siendo superior en términos de mejores valores de CME, AIC, BIC, -2 Log likelihood y Pseudo-R²; que la función gamma incompleta y que el modelo monofásico, por lo que la caracterización de la curva de lactancia de cabras F₁ (Alpina x Nubia) es mejor modelada con dicho modelo.

REFERENCIAS

- De Boer J. A., Weller J. I., Gipson T. A. and Grossman M. 1989. Multiphasic analysis of milk and fat yield curves of Israeli Holsteins. *J. Dairy Sci.* 72: 2143-2152.
- Epaphras A., Karimuribo E. D. and Msellem S. N. 2004. Effect of season and parity on lactation of crossbred Ayrshire cows reared under coastal tropical climate in Tanzania. *Livest Res for Rural Dev* 16 (6) *Art. #42*.
- Faro L. E. e Albuquerque L. G. 1999. Comparação dos ajustes da curva de lactação de vacas Caracu de acordo com três funções matemáticas. In: XXXIII Reuniao anual da sociedade brasileira de zootecnia. pp. 15-17.
- Faro L. E. e Albuquerque L. G. 2002. Comparação de alguns modelos matemáticos para o ajuste às curvas de lactação individuais de vacas da raça Caracu. *Arq. Bras. Med. Vet. Zootec.* 54 (3):S0102-S0109.
- García E. 1973. Modificaciones al sistema de clasificación climática de Köppen. 2ª Ed. UNAM. México, D. F. 246 p.
- Gipson T. A. and Grossman M. 1989. Diphasic analysis of lactation curves in dairy goats. *J. Dairy Sci.* 72:1035-1044.

- Gipson T. A. and Grossman M. 1990. Lactation curves in dairy goats: a review. *Small Rum. Res.* 3: 383-396.
- Gipson T. A., Fernando R. L. and Grossman M. 1990. Effects of smoothing data on estimation of parameters for multiphasic lactation curves of dairy goats. *Livest. Prod. Sci.* 24: 205-221.
- Gonçalves T. M., Gomes A. I., Fonseca F. R., Garcia P. I. 2002. Curvas de lactação em rebanhos da raça Holandesa no Estado de Minas Gerais. Escolha do modelo de melhor ajuste. *R. Bras. Zootec.* 31(4):1689-1694.
- Grossman M. and Koops W. J. 1988. Multiphasic analysis of lactation curves in dairy cattle. *J. Dairy Sci.* 71:1598.
- Juneja V. K., Valenzuela M. M., Huang L., Gumudavelli V., Subbiah J., Thippareddi H. 2006. Modelling the effect of temperature on growth of Salmonella in chicken. doi:10.1016/j.fm.2006.08.004
- Lombaard C. S. 2006. Hierarchical bayesian modelling for the analysis of the lactation of dairy animals. Ph. D. Thesis. Faculty of Economic and Management Sciences, Department of Mathematical Statistics, University of the Free State Bloemfontein.
- Mellado M., Rodríguez S., López R., Rodríguez A. y García J. E. 2004. Relación entre la producción y composición de la leche y el perfil sanguíneo de las cabras al inicio de lactancia en agostaderos. *Memorias de la XIX Reunión Nacional sobre Caprinocultura.* pp. 276-280.
- Montaldo H., Almanza A., Juárez A. 1997. Genetic group, age and season effects on lactation curve shape in goats. *Small Rum Res* 24: 195-202.
- Olori V. E., Brotherstone S., Hill W. G. and McGuirk B. J. 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.* 58:55-63.
- Rekaya R., Weigel K. A., Gianola D. 2001. Hierarchical nonlinear model for persistency of milk yield in the first three lactations of Holsteins. *Livest. Prod. Sci.* 68: 181-187.
- Rekik B. and Ben A. G. 2004. Factors affecting the occurrence of atypical lactations for Holstein–Friesian cows. *Livest. Prod. Sci.* 87: 245-250.
- Ruvuna F., Kogi J. K., Taylor J. F. and Mkuu S. M. 1995. Lactation curves among crosses of Galla and East African with Toggenburg and Anglo Nubian goats. *Small Rum. Res.* 16: 1-6.
- SAS. 2002. SAS User's Guide: Statistics. Institute Inc. Cary, NC, USA
- Schabenberger O., 2008. Nonlinear Regression in SAS. University of California, Los Angeles, SAS Library. http://www.ats.ucla.edu/stat/SAS/library/SASNLin_os.htm.
- Scott T. A., Yandell B., Zepeda L., Shaver R. D., and Smith T. R. 1996. Use of lactation curves for analysis of milk production data. *J. Dairy Sci.* 79: 1885-1894.
- Sherchand L., McNew R. W., Kellogg D. W. and Johnson Z. B. 1995. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *J. Dairy Sci.* 78: 2507-2513.
- Tozer P. R. and Huffaker R. G. 1999. Mathematical equations to describe lactation curves for Holstein-Friesian cows in New South Wales. *Aust. J. Agric. Res.* 50:431-440.
- Weigel K. A., Craig B. A., Bidwell T. R., and Bates D. M. 1992. Comparison of alternative diphasic lactation curve models under bovine somatotropin administration. *J Dairy Sci* 75: 580-589.

Algunos Aspectos Estadísticos sobre un Estudio de Salud Nutricional y Salud Dental en Niños de Enseñanza Media en el Estado de Veracruz

Luisa Otilia Navarro Ortiz¹, Luis Cruz-Kuri², Paula Yadira Landa Cano³, Julia Leticia Beltrán Martínez³, María Magdalena Palmeros Rodríguez³

¹Facultad de Odontología, U.V. lonol24@hotmail.com

²Instituto de Ciencias Básicas, U.V. kruz1111@yahoo.com.mx

³Centro de Salud “Dr. José A. Maraboto”

1. Introducción

El presente reporte se ubica dentro de un estudio más amplio realizado por algunos de nosotros. En dicho proceso se registraron variables sobre salud bucal, estado nutricional, costumbres alimenticias y de higiene, entre otras. El estudio incluye alumnos de escuelas primarias, así como de escuelas secundarias. Por lo que corresponde a las fechas en que se realizaron las encuestas, estas se realizaron desde 1999 y continuaron hasta el presente año de 2009.

Iniciamos acudiendo a las escuelas a revisar la boca, a pesar y medir la talla a los niños dentro del programa *Salud Escolar* ahora *Escuela Saludable*. Al término de la tarea, obteníamos, entre otras variables, el *Índice de Masa Corporal*, lo comparábamos en la Norma Oficial Mexicana 008 SSA2 1993 y coincidía que, los niños que padecían caries estaban desnutridos o con sobre peso. Nos inquietó esta vinculación y quisimos investigar cuales eran las causas que motivaban esta relación. Formulamos un cuestionario de costumbres alimenticias y de higiene. Levantamos las encuestas verticales en 20 escuelas de niveles primario y secundario de la ciudad de Xalapa, San Andrés Tlanahuayocan y Xico, todas del estado de Veracruz. Decidimos descartar a los niños de preescolar porque estos tienen más control en la alimentación y aseo bucal dentro de la escuela.

2. Procesamiento de los Datos y Algunos Resultados

Para los datos relevantes, se incluyeron las variables de edad, peso, talla, género, grado escolar, índice de masa corporal, peso ideal y un factor de peso, con este último, para determinar el estado nutricional del alumno; asimismo, se registró el estado dental, entre otras variables. Toda esta información fue integrada en hojas de cálculo para ser analizadas mediante un programa de cómputo estadístico. El programa utilizado fue el SPSS versión 17.0, con el cual se pueden hacer procesamientos numéricos tanto de tipo univariado, como de tipo multivariado, así como salidas gráficas. Con la capacidad de selección de casos del programa, se realizaron análisis condicionales, utilizándose variables tales como la de género, estado dental, estado nutricional, edad, entre otras. Por limitaciones de espacio, para el presente reporte, se omiten los detalles de las instrucciones específicas que generan los archivos condicionados; asimismo, se omiten detalles sobre los análisis de tipo multivariado que pueden realizarse. En una de las secciones, se presentan resultados para las escuelas primarias y en otra de las secciones los correspondientes para las escuelas secundarias.

3. Escuelas Primarias

Para la presente parte se consideran únicamente las escuelas primarias. El total de alumnos encuestados fue de 780 estudiantes. Las primarias consideradas fueron cuatro. Por razones de espacio, en el presente reporte solo se han seleccionado algunos años, los cuales son los que se indican enseguida. Año de 1999 con 205 alumnos (representando un porcentaje de 26.3%); año de 2004 con 100 alumnos (12.8%); año de 2006 con 127 alumnos (16.3%); año de 2007 con 240 alumnos (30.7%) y año de 2008 con 108 alumnos (13.8%). En lo correspondiente a género de los alumnos, 420 fueron de sexo masculino (53.8%) y 360 de género femenino (46.2%).

En lo referente a su estado nutricional, según su índice de masa corporal un 68.4% eran clasificados con desnutrición moderada, desnutrición leve o desnutrición mayor; el 26.2% de los alumnos cayeron en la categoría de estado nutricional normal; cerca del 2% fueron clasificados como obesos y otro 3.5% como con sobrepeso.

En el sistema mexicano los grados de las escuelas primarias van desde el primero hasta el sexto. Para nuestro reporte en el primer grado se encontraron aproximadamente el 10% de todos los alumnos encuestados; en el segundo grado casi el 11.5 %; en el tercer grado casi el 18%; en el cuarto grado aproximadamente el 20% en el quinto grado aproximadamente el 17.5% y en el sexto grado casi el 24%. Ver. Fig. 1.

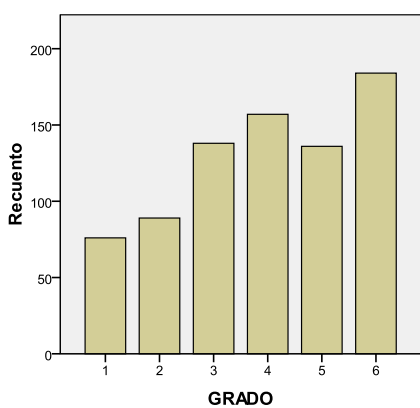


Fig. 1. Distribución de alumnos de escuelas primarias por grado escolar.

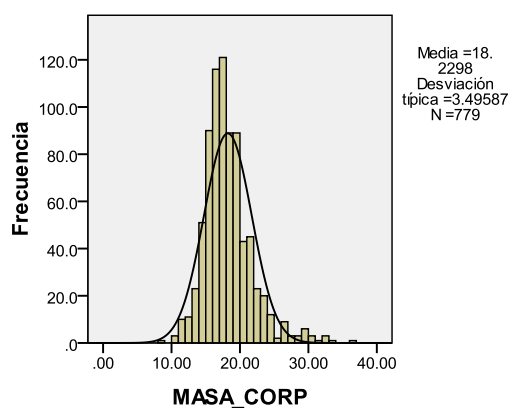


Fig. 2. Distribución de índice de masa corporal para alumnos de primaria.

Tal como ya se hizo notar, más del 68% de los alumnos encuestados cayeron en categorías de algún tipo de desnutrición. En lo que corresponde a su estado de salud dental casi el 75% de los niños estudiados presentaban caries y únicamente el 25% estaban libres de caries. El promedio para el índice de masa corporal fue de 18.23, con una desviación estándar de 3.496. Ver Fig. 2 arriba.

De la distribución de frecuencias correspondiente, se encontró que 470 alumnos tenían índice de masa corporal inferior a 18.4, con lo cual caen en las categorías de desnutrición leve, moderada o severa, siendo el índice promedio de 16.12 y de estos el 75.1%

presentaron problemas dentales de caries. En comparación, aquellos alumnos con índice corporal de masa mayor igual que 18.4 corresponden a 309 de los cuales el 74.4% presentan caries, lo cual dice que el estado nutricional está ligeramente asociado con el estado de salud dental. Parte de la misma información, pero en forma gráfica, aparece en la Fig. 2.

Del total de los 780 alumnos, 584 presentaron caries. De estos, el 45% corresponden a mujeres y el 55% a varones. El 60.5% de todos los alumnos con caries presentaban un índice de masa corporal inferior al 18.4, con lo cual se puede decir que más de la mitad de los alumnos con caries caían en un estado de desnutrición. En comparación, al analizar las correspondientes estadísticas para alumnos sin caries los resultados son los que se comentan enseguida.

Se tienen ahora 195 casos de alumnos que no presentan caries, teniéndose aproximadamente cantidades iguales para los dos géneros. En lo referente a su índice de masa corporal el 60% están por debajo de 18.4, lo cual nuevamente indica que más de la mitad de estos alumnos (ahora sin caries también tienen varios grados de desnutrición). Al comparar la cifra de 60.5% de desnutridos y con caries contra el 60.0% de desnutridos pero sin caries, la diferencia de medio punto porcentual es marginal.

4. Escuelas secundarias

Los análisis que siguen están basados en encuestas realizadas a 655 alumnos en distintas épocas desde 1999 hasta el presente.

De los alumnos de secundaria 336 corresponden a mujeres y 319 a varones, con lo cual los respectivos porcentajes son de 51.2% y 48.8% aproximadamente; el 35.5% tiene un grado de desnutrición que va desde leve hasta severo; aproximadamente el 56.6% tienen grados de nutrición que pueden considerarse como normal. Solo se incluyen dos escuelas secundarias; a saber, las secundarias Zapata con 439 alumnos encuestados y la secundaria Zárata con 216 alumnos encuestados. La edad que ocurre con más frecuencia es la de 13 años y le sigue la de 12 años en tercer lugar la de 14 en cuarto la de 15 en quinto la de 16 le sigue la de 17, hay muy pocos alumnos con 18 años y un número todavía más bajo con 11 años de edad.

Para estos alumnos únicamente el 30.3% tienen un índice de masa corporal inferior a 18.38, lo cual está indicando que menos de la tercera parte de todos los alumnos tienen problemas de desnutrición. Compárese esta cifra con la correspondiente en la primaria donde el porcentaje de desnutridos es de más del doble que la de correspondiente categoría en la secundaria.

Algunos alumnos presentan obesidad extrema. Por ejemplo, el estudiante número 564 (según el número progresivo asignado en nuestro archivo) tiene un índice corporal de masa de 72.84; al ver cual era su peso este fue de 63 kilogramos y su estatura de 93 cm. El caso número 540 corresponde a un alumno con estatura de 99 cm. y con un peso de 43 kilogramos, dando lugar a un índice de masa corporal de 43.87.

Análisis por Género para Alumnos de Secundaria

Tabla 1. Frecuencias. Estado dental para varones *con* algún grado de desnutrición.

| DENTAL | | | | |
|---------|-------|-------|------------|-------------------|
| | | Frec. | Porcentaje | Porcentaje válido |
| Válidos | NO | 71 | 39.2 | 39.2 |
| | SI | 110 | 60.8 | 60.8 |
| | Total | 181 | 100.0 | 100.0 |

Tabla 2. Frecuencias. Estado dental en varones *sin* desnutrición.

| DENTAL | | | | |
|---------|-------|-------|------------|-------------------|
| | | Frec. | Porcentaje | Porcentaje válido |
| Válidos | NO | 74 | 54.4 | 54.4 |
| | SI | 62 | 45.6 | 45.6 |
| | Total | 136 | 100.0 | 100.0 |

Puede verse que la mayoría de los varones de los varones con algún grado de desnutrición tienen caries (60.8%). En contraste, se tiene un porcentaje menor (45.6%) de varones con caries cuando no tienen problemas de desnutrición.

Tabla 3. Frecuencias. Estado dental de mujeres *con* desnutrición.

| DENTAL | | | | |
|---------|-------|-------|------------|-------------------|
| | | Frec. | Porcentaje | Porcentaje válido |
| Válidos | NO | 47 | 35.9 | 35.9 |
| | SI | 84 | 64.1 | 64.1 |
| | Total | 131 | 100.0 | 100.0 |

Tabla 4. Frecuencias. Estado dental de mujeres *sin* desnutrición.

| DENTAL | | | | |
|---------|-------|-------|------------|-------------------|
| | | Frec. | Porcentaje | Porcentaje válido |
| Válidos | NO | 56 | 27.3 | 27.3 |
| | SI | 149 | 72.7 | 72.7 |
| | Total | 205 | 100.0 | 100.0 |

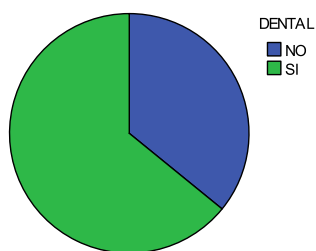


Fig. 3. Estado dental de mujeres *con* desnutrición. *Nota:* la codificación SI corresponde a presencia de caries.

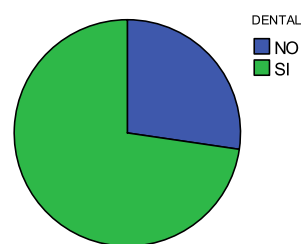


Fig. 4. Estado dental en mujeres *sin* desnutrición. *Nota:* la codificación SI corresponde a presencia de caries.

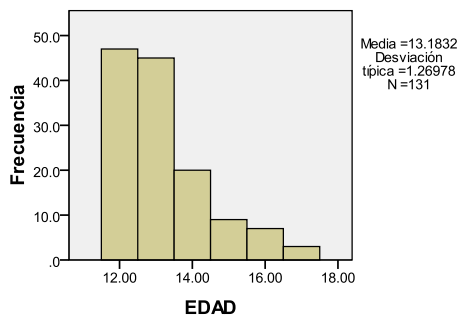


Fig. 5. Distribución de edades para mujeres *con* desnutrición.

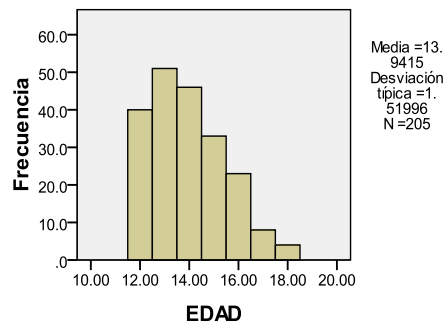


Fig. 6. Distribución de edades para mujeres *sin* desnutrición.

Puede notarse, al comparar las distribuciones de edades en los dos grupos (Fig. 5 y Fig. 6) que las mujeres clasificadas con algún tipo de desnutrición, tienen en el promedio una edad más baja que aquellas que no presentan desnutrición. Esto, sugiere la posibilidad de que las mujeres con algún tipo de desnutrición estarán sujetas durante más tiempo a condiciones adversas que pueden favorecer la presencia de caries. Por otra parte, al analizar a los alumnos por grupos de edad, se encontró que el grupo de 17 años de edad presentaba el factor de peso más grande, y que el que le seguía en orden de magnitud correspondía al grupo de los de 12 años de edad; los grupos de 13, 14, 15 y 16 años de edad presentaron un factor de peso sustancialmente por debajo de los de 17 y de 12 años de edad.

5. Comentarios Adicionales

Después de haber entrevistado, revisado cavidades bucales, pesado y medido, a más de 3207 niños de edad escolar, podemos afirmar a través de los registros, y de los análisis estadísticos multivariados aplicables, que el estado nutricional de los niños si se ve afectado por el descuido de la salud bucal y viceversa, el estado nutricional de alguna manera está relacionado con el estado dental, entre otros factores. Por consideraciones de espacio, para el presente trabajo, solo se han reportado algunos de los análisis estadísticos correspondientes a una parte de la información recabada.

REFERENCIAS

- [1] Colton, Theodore (1974) *Statistics in Medicine*. Little Brown & Co.
- [2] Daniel, Wayne W (2009) *Biostatistics: A Foundation for Analysis in the Health Sciences + Spss Student Version 16.0*. John Wiley & Sons Inc.

Algunas Estadísticas de la Precipitación en el Puerto de Veracruz

Antonio Luna Díaz Peón¹, Sergio Hernández González²

¹Universidad Veracruzana. diazpeon@yahoo.com.mx

²Universidad Veracruzana. s_her_gon@yahoo.es

Resumen

Veracruz es el estado del país que presenta mayor frecuencia de inundaciones. Además, en Veracruz se concentra una importante cantidad del caudal de los ríos del país, el mayor volumen de agua superficial de México, es un estado vulnerable a los ciclones tropicales que en ocasiones acarrearán grandes cantidades de agua, entre otros efectos. El Puerto de Veracruz en cuestiones meteorológicas es de gran importancia pues se han medido estas variables desde los siglos XVII que se tenga conocimiento, por ello es de destacar con sus estadísticas el comportamiento de la precipitación a lo largo del tiempo con que se cuenta registro que es desde 1917.

1. Introducción

En México la precipitación pluvial causa anualmente inundaciones que ocasionan destrucción de vías de comunicación y viviendas, afectaciones en las actividades económicas, y, lo peor, pérdida de vidas humanas. Conocer y aprender más sobre el comportamiento de la precipitación se convierte en una prioridad para las actividades de protección civil.

Veracruz es el estado de México que presenta mayor frecuencia de inundaciones con 3.5 eventos por año (Ortiz et al., 1991) y concentra el mayor volumen de agua superficial con más del 30% del caudal de ríos del país (Luna, 2007). El Puerto de Veracruz de igual forma año con año presenta inundaciones que en algunos puntos hoy en día se reflejan de gran importancia al quedar incomunicadas colonias por períodos de tiempo.

2. El Puerto de Veracruz

Se encuentra ubicado en la zona centro del Estado, región sotavento, en las coordenadas 19° 12' latitud norte y 96° 08' longitud oeste a una altura de 10 metros sobre el nivel del mar. Limita al norte con La Antigua, al este con Boca del Río y el Golfo de México y al sur con Manlio Fabio Altamirano y Medellín.



Figura 1. División política y regionalizada del Estado de Veracruz.

El Puerto de Veracruz según información de la Comisión Nacional del Agua registra una precipitación acumulada media anual de entre 1600 a 1800 mm (fig. 2). La clasificación climática de Koeppen modificada por García (1981) indica que la zona de estudio presenta la categoría de clima cálido y húmedo de los que presentan lluvias en verano (Aw2).

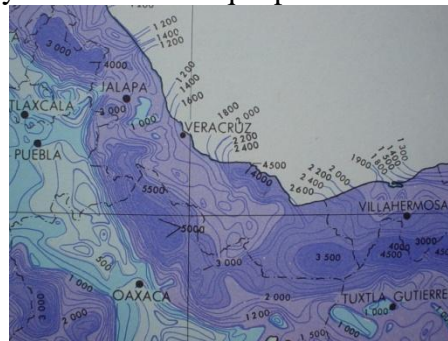


Figura 2. Precipitación acumulada anual. Fuente: Comisión Nacional del Agua (1992).

3. Los datos

El observatorio meteorológico de Veracruz (dependiente de la GRGC) cuenta con datos registrados desde 1917, aunque se sabe que durante el período de 1790 a 1803, el capitán del puerto don Bernardo de Orta, reunió una serie de datos que utilizó el barón Alejandro de Humboldt para efectuar un estudio del clima en la Nueva España y que constituyeron el registro de observaciones más largo que se ha encontrado en América.

Los datos que actualmente se tienen de registro y que son en forma diaria muestran que coinciden con los análisis realizados en 1992, el promedio de acumulado anual de precipitación 1917 al 2008 es de 1,703.7 (fig. 3). Claramente se muestra lo también ya conocido como canícula o sequía intraestival (es el suceso conocido cuando en la temporada de lluvias que es entre mayo y octubre, existe una disminución del acumulado que para nuestra área de estudio es en agosto).

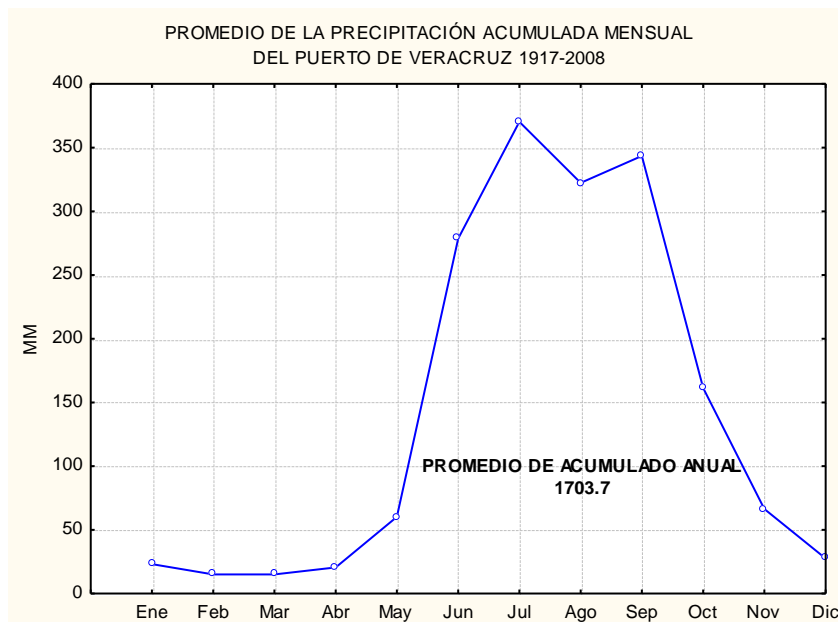


Figura 3. Promedio de la precipitación acumulada mensual y promedio acumulado anual del Puerto de Veracruz 1917-2008.

Es de destacar que los datos existentes son desde el primero de enero de 1917 hasta el 31 de diciembre de 2008, sumando con ello 33,530 observaciones de precipitación. En todo este periodo de análisis ha existido variabilidad de tal forma que incluso en 1955 el acumulado de la precipitación fue de 2,649.3 mm y en 1943 se registraron 731.2 mm.

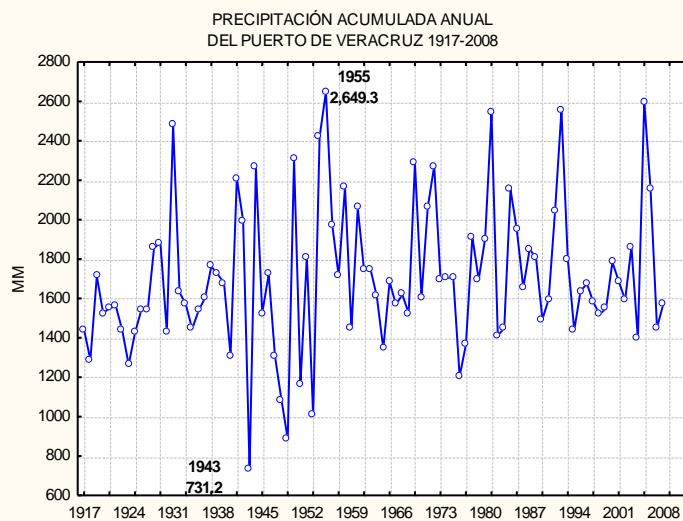


Figura 4. Precipitación acumulada anual 1917-2008 del Puerto de Veracruz.

Sin embargo la mayor concentración de la precipitación es alrededor del promedio, una forma más clara de visualizarlo es mediante el gráfico de “cajas y alambres” (fig.5), ya que se puede observar dicha variabilidad.

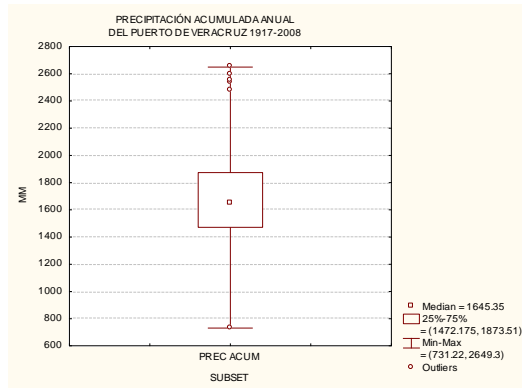


Figura 5. Precipitación acumulada anual 1917-2008 del Puerto de Veracruz mediante el gráfico de “cajas y alambres”.

4. Algunas estadísticas de la precipitación en el Puerto de Veracruz

Debido a los problemas que a nivel mundial se plantean por efectos del Cambio Climático Global, en Veracruz se cuenta con un Plan Veracruzano ante el Cambio Climático (PVCC) el cual indica que para el Golfo de México según Tejeda (2009), el escenario de cambio climático global más probable es aquél que, en resumen, prevé que a lo largo de este siglo los eventos extremos del clima (sequías, lluvias abundantes, huracanes más intensos) se acentuarán. A este escenario se suman dos fenómenos mayores: el cambio en las condiciones térmicas de la corriente del Golfo, y el aumento en el nivel de los mares relacionado con el deshielo de los glaciares continentales y la dilatación térmica, de acuerdo a los resultados difundidos por el Panel Intergubernamental de Cambio Climático (IPCC, por sus siglas en inglés) en su cuarto reporte de evaluación.

Para nuestro caso el pronóstico tiene mucha variabilidad ya que algunos modelos que se utilizan mencionan que aumentará la precipitación hacia mediados y finales del presente siglo y otros que disminuirá, pero para entender esto veamos que es lo que realmente ha sucedido, si graficáramos cada una de las series de tiempo de cada año el resultado sería el siguiente (fig.6).

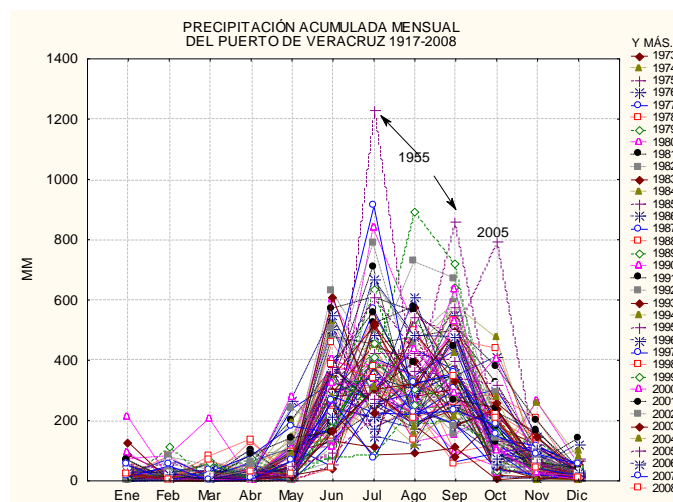


Figura 6. Precipitación acumulada mensual 1917-2008 del Puerto de Veracruz

Existe una gran cantidad de información, destacando que año con año el comportamiento de la precipitación se conserva, además existen extremos altos por ejemplo lo sucedido en 1955, ya que en dos meses (julio y septiembre) reflejan los históricos (fig. 6), en lo que respecta a julio se desconoce que fenómenos fueron los que influyeron ya que no se cuenta con dicha información que lo muestre, sin embargo para el mes de septiembre se atribuyen dichas precipitaciones a tres ciclones tropicales que afectaron a Veracruz y que la información existe y coincide con cada uno de los fenómenos, veamos cada uno de ellos, el primero de ellos “**Gladis**” (fig. 7) que alcanzó la categoría de huracán antes de impactar el extremo norte el día 6 donde ocasionó que el río Pánuco se desbordara pérdidas materiales, de cultivo y humanas (Diario de Xalapa, 1955). El segundo fue el ciclón tropical “**Hilda**” (fig. 7) 13 días después tocando tierra de igual forma en el norte de Veracruz donde nuevamente afectó, pero antes provocó en la cuenca del Papaloapan inundaciones considerables. Y el tercero y más importante para nuestro estudio de caso fue el ciclón tropical “**Janet**” (fig. 7), ya que 10 días después (29 de septiembre) impactó su centro como categoría de huracán a la altura de Emilio Carranza, sin embargo sus bandas nubosas cubrieron el Puerto de Veracruz provocando una precipitación el día 28 de 340.1 mm y el día 29 148.5 mm (488.6 mm en 48 horas, fig. 8). La precipitación ocurrida provocó corrientes caudalosas afectando casas y sembradíos, además de animales muertos. Se mencionó que 10 personas se ahogaron, aunque el dato no se precisó, estimando que pudieron haber sido más, es de destacar que cuando este sistema afectó a Veracruz algunas zonas aun permanecían inundadas. (Diario de Xalapa, 1955).

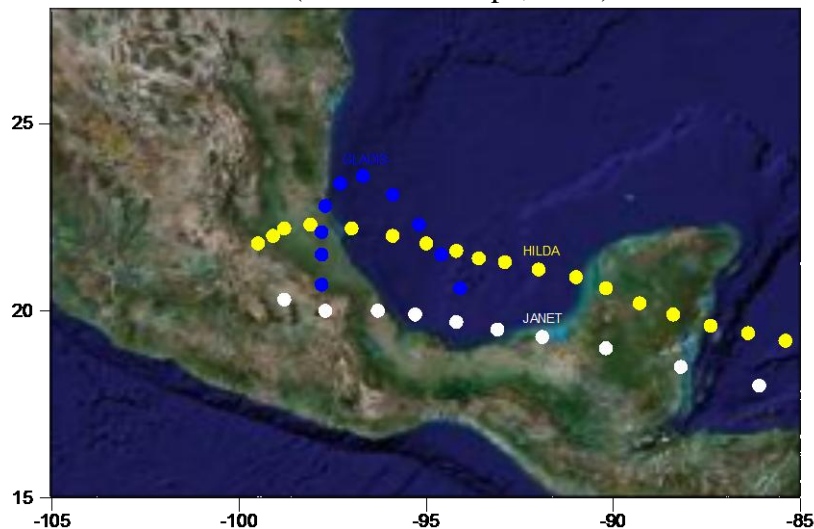


Figura 7. Ciclones tropicales que impactaron costas veracruzanas durante el mes de septiembre de 1955.

Denotemos que los tres ciclones tropicales ocurrieron sólo en el mes de septiembre del año 1955 y por ello el máximo histórico de precipitación mensual ocurrió en ese mes (fig.8), incluso algunos medios de comunicación de la época denotaron el problema que ocasionaron y que hasta la fecha es también el máximo histórico de frecuencias de ciclones tropicales que han impactado a Veracruz en un mes. Ya que en toda la temporada fue en el 2005 con 4 fenómenos ocurridos.

PRECIPITACIÓN DIARIA DEL PUERTO DE VERACRUZ
SEPTIEMBRE DE 1955

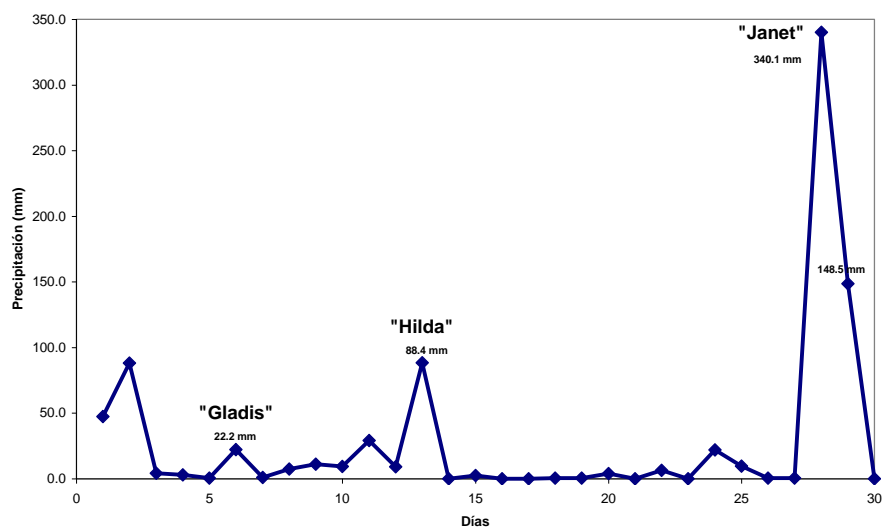


Figura 8. Precipitación durante el mes de septiembre de 1955.

En el año 2005 durante el mes de octubre es que se registró el máximo mensual de precipitación en el Puerto de Veracruz con 793.1 mm (fig.6), destacando los días 4 y 5 cuando ingreso el ciclón tropical “Stan” al sur de la entidad veracruzana a la altura de los Tuxtlas dejando en esos dos días un acumulado de precipitación de 435.8 mm, la cantidad registrada fue ligeramente inferior a la de 1955, sin embargo es de destacar dos cosas importantes: la primera que los daños materiales fueron cuantiosos y en algunos casos incalculables, también hubo pérdidas materiales y en la agricultura, pero lo segundo y más importante es que no se tiene registro de pérdidas de vidas humanas ocasionadas directamente por el sistema ciclónico. El cambio de uso de suelo y el aumento en la contaminación ambiental pudiera acarrear en un futuro no muy lejano problemas mas fuertes de inundaciones, por lo que el presente trabajo aun se le puede incrementar y mejorar, con la posibilidad de que sirva de ayuda a las autoridades competentes. El Cambio Climático Global es evidente y somos nosotros los humanos el principal factor que lo ocasiona, sin embargo somos nosotros que lo podemos evitar apoyados de las recomendaciones ya bastas existentes.

REFERENCIAS

- CNA, Comisión Nacional del Agua (1992), Isoyetas normales de la República Mexicana 1931-1990, México.
- García, E. (1981). *Modificaciones al sistema de clasificación climática de KOPPEN*. México.
- Luna, A. (2007). *Modelo Espacio-Temporal de la precipitación estudio de caso: Cuenca del Papaloapan*. Tesis de Maestría. Universidad Veracruzana.
- Ortiz, M. (1991). *Mapa de riesgos naturales, Atlas nacional de México*. Instituto de Geografía, UNAM.
- Tejeda, A. Compilador. (2009). Programa veracruzano ante el cambio climático. Gobierno del Estado de Veracruz.

Propuesta, y Ejemplo, de un Procedimiento Multivariante para Integrar la Metodología Cualitativa (con Material de Texto Libre) y la Metodología Cuantitativa (Estadística Multivariante)

José David Urchaga Litago
Univ. Pontificia de Salamanca. España. urchaga.upsa.es

Resumen

En el presente trabajo se propone un marco que integre la metodología cualitativa (inductiva) y cuantitativa (deductiva), de tal forma que se pueda trabajar un material libre (cualitativo, como son los textos) con técnicas estadísticas multivariantes.

1. Los enfoques cualitativo y cuantitativo

Tradicionalmente, en las ciencias sociales, se han visto como opuestas la metodología cualitativa y cuantitativa. Son muchos los autores que remarcan sus diferencias, y en el mejor de los casos las consideran como complementarias (Delgado y Gutiérrez, 1995), pero son pocas las propuestas de integración. Newman y Benz (1998) en su obra *Qualitative-Quantitative research methology: exploring the Interactive continuum* plantearon un modelo inclusivo hasta ahora poco conocido en el ámbito Iberoamericano (Berganza y Sánchez 2005; Kerlinguer y Howard, 2002). Newman y Benz proponen un continuo circular en la metodología de investigación cuantitativa y cualitativa. Proponen las siguientes fases:

| | | | | | |
|--------------|-----------------|-------------|----------------|-------------|-----------------|
| Cuantitativo | I.Teoría | 2.Hipótesis | 3.Datos | 4.Análisis | 5.Conclusiones |
| Cualitativo | A.Datos | B.Análisis | C.Conclusiones | D.Hipótesis | E.Teoría |

Por lo que ambas tienen en común que integran los mismos elementos (teorías, hipótesis, datos...), por lo que es posible establecer un continuo, en el sentido de poder contrastar las teorías propuestas desde el método cualitativo. El método cualitativo sería inductivo, parte de la realidad para *construir teorías*, mientras que el cuantitativo sería deductivo y *testaría las teorías* desde las que comienza.

En ciertos ambientes académicos –especialmente en los de humanidades– se consideran ambas posturas irreconciliables, ya que, desde el método cualitativo, se afirma que los métodos cuantitativos en su obsesión por medir suelen dejar de lado aspectos muy importantes de lo humano (y en cierto modo tienen razón, ya que sólo miden aquello que pueden medir, no todo lo que les gustaría medir). Por otro lado, desde el método cuantitativo se les acusa que el método inductivo genera teorías poco falsables (recuérdese las críticas de Popper) y generalmente poco parsimoniosas y muy subjetivas, que raramente se contrastan.

2. Problema a resolver

El punto central de nuestra propuesta es analizar material cualitativo con estadística multivariante, de tal forma que se puedan crear y testar teorías. Un buen intento de metodología es el análisis de contenido (Urchaga, 2007), que se inserta dentro del método cuantitativo, ya debe crear unas categorías de análisis que se derivan de las teorías. Otra

aproximación, desde el método cualitativo, son los análisis realizados con programas informáticos tales como el Nudist, por el cual las categorías se van creando a partir de los datos. Estos dos tipos de análisis adolecen de dos limitaciones, el primero es que inevitablemente al categorizar el material, éste, inevitablemente, pierde su referencia original; en segundo lugar, las estadísticas que se realizan suelen ser descriptivas, y relacionando muy pocas variables, y esto suele ser debido al desconocimiento de cómo adaptar las bases de datos (suelen ser enormes) para poder realizar estadísticas multivariantes (que en muchos casos también son desconocidas).

3. Propuesta

El Dr. Urchaga, en la Universidad Pontificia de Salamanca, ha realizado varias investigaciones que partiendo de material cualitativo, principalmente textos (por ejemplo, obras de literatura, cuestionarios de respuesta abierta, noticias de periódicos) ha empleado técnicas multivariantes (en especial, el análisis de cluster) poco conocidas en el ámbito de las ciencias de la comunicación, y que en ámbito de las ciencias humanas y sociales son empleadas casi en exclusividad con variables no cualitativas.

Un primer trabajo (Urchaga, 2005) replicó la investigación de Deconchy (1967) sobre la evolución de la idea de Dios con 2761 escolares (entre 7 y 18 años). El cuestionario era del tipo de “respuesta y asociación libre”, por lo que cada participante podía dar respuestas únicas. Cada persona debía escribir cinco palabras que asociaran a una palabra estímulo (para nuestro caso sólo atenderemos las respuestas dadas a la palabra-estímulo Dios). El formato del cuestionario era el siguiente (anotamos en cursiva las respuestas dadas por una niña de 10 años:

| | | | | | |
|-------|-------------|---------------|---------------|---------------|---------------|
| PADRE | <i>Ropa</i> | <i>Amor</i> | <i>Ayuda</i> | <i>Dinero</i> | <i>Enseña</i> |
| MADRE | <i>Amor</i> | <i>Ayuda</i> | <i>Enseña</i> | <i>Regala</i> | <i>Cuida</i> |
| DIOS | <i>Ama</i> | <i>Enseña</i> | <i>Cuida</i> | <i>Paz</i> | <i>Ama</i> |

Deconchy había categorizado las respuestas, pero debido a que dichas categorías eran más teológicas que psicológicas se exploró si era posible otra forma de analizar dichas respuestas. La hipótesis que se quería testar era si la idea de Dios variaba en función de la edad y del sexo.

La primera base de datos tenía un formato de 2761 casos (filas) por las cinco respuestas, más el grupo de edad y el sexo. La primera transformación consistió en separar esta base en otras en función del grupo de edad (8 a 18 años) y del sexo (hombre/mujer), por lo que se pasó a 22 bases de datos. Cada una de estas bases se transformó en una nueva donde en una única variable se ponían todas las respuestas dadas por las personas de dicha base, así se pasaba de una base de $n \times 5$ (respuestas), a una de $5n \times 1$. Posteriormente, se solicitaba al programa estadístico SPSS que nos mostrará una tabla de frecuencias de dicha variable, de tal forma que estuvieran ordenadas las respuestas de mayor a menor frecuencia. En esta tabla de frecuencias, los porcentajes (los vamos a denominar fr) son sobre el total de respuestas, no sobre el de sujetos, por lo que para hacernos una idea aproximada del porcentaje de sujetos que daban un tipo de respuesta había que multiplicar el porcentaje por cinco (esta es una buena aproximación, ya que normalmente las respuestas no se solían

repetir). Así se obtenía para cada grupo de edad y sexo qué palabras eran las más frecuentes. Se recogieron 1,705 diferentes respuestas, pero sólo nos interesaban aquellas que fueran representativas de los grupos, es decir, aquellas que los definieran (por ser de alta frecuencia; alta *fr* grupal) y que los diferenciaran de otros grupos (ya que estuvieran más presentes en unos que en otros; alta variabilidad intergrupo). Se optó por trabajar con aquellas 10 más representativas de cada grupo, ya que tras un estudio previo se comprobó que con este número se podían realizar análisis de conglomerados coherentes con los modelos teóricos. Como se puede apreciar en la siguiente tabla, no todas las palabra-respuesta (Pr) son igual de representativas de los grupos, y son muy pocas las palabras que se repiten en los grupos (% sujetos que escriben la Pr):

| % sujetos que escriben la Pr | Nº Pr | % Pr | Pr y edad donde la <i>fr</i> es máxima |
|------------------------------|-------|------|--|
| 53,45 | 1 | 0,1 | Bueno, 9/10 |
| 41,61 | 1 | 0,1 | Padre, 14/15 |
| 31,62 | 1 | 0,1 | Iglesia, 16/17 |
| 31,58 | 2 | 0,1 | Poderoso, Amable, 11/12 |
| 30,95 | 1 | 0,1 | Religión, 15/16 |
| 29,20 | 1 | 0,1 | Amor, 13/14 |
| 27,01 | 1 | 0,1 | Cielo, 7/8 |
| 20 -24,9 | 4 | 0,4 | Creador y grande, 11/12; Amigo, 13/14; fe, 17/18; |
| 15 -19,9 | 4 | 0,2 | N.C. y Jesús, 7/8; nada y misa, 17/18 |
| 10 -14,9 | 8 | 0,5 | Guapo y paz 8/9; cariñoso y alto 9/10; señor,11/12; salvador, 12/13; ayuda,15/16; creencia, 17/18; |
| 3 - 9,9 | 112 | 10,4 | |
| 0 - 2,9 | 1158 | 92,1 | |
| Total | 1720 | 100 | |

Como se puede observar, hay 8 palabras que han sido repetidas al menos por el 25% de los sujetos, por lo que representan a gran parte de la muestra. La mayoría de las palabras sólo han sido escritas por uno o dos sujetos, por lo que no son representativas de los grupos.

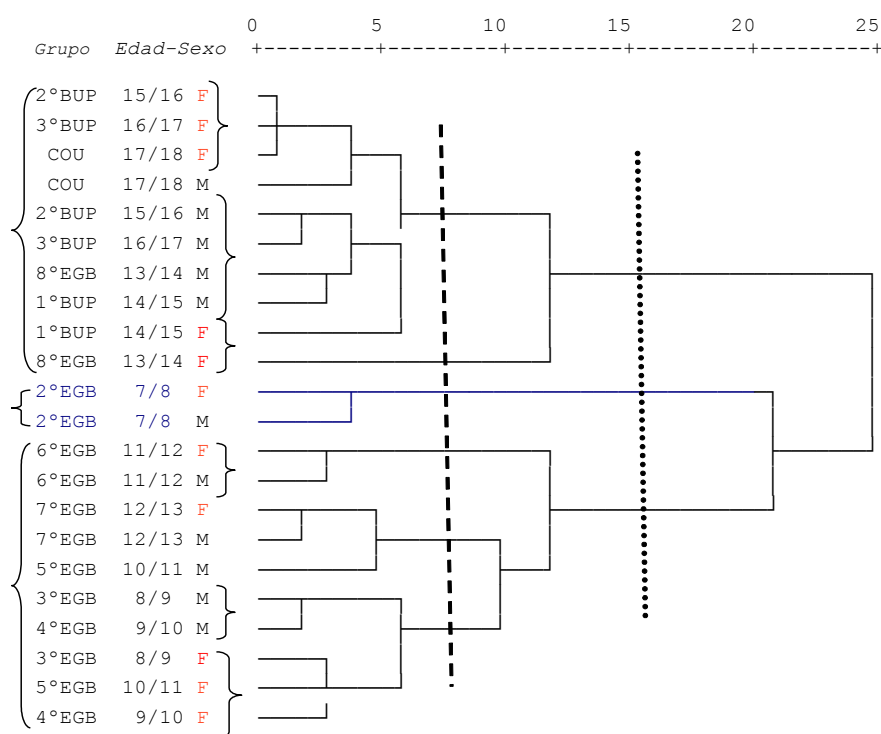
A partir de estos datos se creó una base de datos definitiva que contenía en filas los nombres de los grupos según sexo y edad (22 grupos, que son nuestros casos), y tantas columnas como palabras diferentes se quisieran utilizar para realizar los conglomerados. En las casillas de datos se ponían como datos los porcentajes (*fr*). Así, la estructura de la base sería la siguiente (no se muestran todas las columnas ni todos los casos):

| Grupo | Bueno | Religión | Iglesia | Padre | N.C. | Poderoso | Fe | Amigo |
|---------|-------|----------|---------|-------|------|----------|------|-------|
| 8/9 F | 10 | 0,42 | 1,67 | 6,25 | 3,96 | 0,63 | 0,21 | 0,63 |
| 9/10 F | 10,69 | 0,35 | 1,03 | 5,52 | 0,52 | 0,69 | 0,17 | 0,69 |
| 17/18 F | 1,60 | 6,14 | 4,54 | 3,56 | 3,68 | 0,25 | 4,40 | 2,70 |
| 8/9 M | 9,05 | 0,42 | 0,63 | 4,63 | 3,37 | 1,90 | 0 | 0,84 |
| 9/10 M | 8,52 | 0,74 | 1,67 | 4,07 | 4,26 | 1,48 | 0,19 | 1,67 |
| 17/18 M | 0,39 | 4,71 | 4,71 | 2,16 | 4,71 | 0,78 | 2,07 | 0,78 |

Este diseño de la base sí permite ahora realizar diferentes análisis estadísticos multivariados. Así, un análisis de conglomerados (con vinculación intergrupos y método de Ward) realizado con las diez palabras-respuesta con mayor frecuencia relativa nos mostró una configuración (más adelante se muestra) que nos permitió describir tres fases evolutivas, pero de diferente forma para los chicos que para las chicas, ya que dentro de cada “fase evolutiva” se conglomeraban antes por sexo que por grupo de edad.

Esto era sorprendente, pero encontramos la respuesta cuando regresamos a comprobar qué palabras eran las más frecuentes en cada edad, y descubrimos una diferencia cualitativa entre los sexos.

Las mujeres utilizan más palabras, y además mantienen una constante en todas las edades – amor- por lo que renuevan los significados sin rechazar todo lo previo (en el siguiente cuadro se observa lo expuesto para tres grupos de edad, en una representación gráfica de frecuencias relativas, donde las palabras se sitúan en una altura en proporción a sus frecuencias relativas dentro de su grupo).



Así, las mujeres a los 8/9 años tienen tres palabras para definir a Dios (Bueno, Padre, Amor), mientras que los chicos sólo una. Además, en el sexo femenino, la palabra Amor está en la base de la construcción y reconstrucción del significado de Dios. De esta forma se pudo establecer algunas matizaciones respecto a los resultados originales de Decocny. Se establecen tres fases: perceptiva (7/8 años cielo, María); Atributiva (9-12 años, Bueno, Amable, Creador); la tercera fase se subdivide en dos: Personal (13-14 en chicas, 13-15 en chicos; padre, Amigo). Interiorizada-institucionalización (hasta los 18, Religión, Duda, Fe).

4. Conclusión, otros ámbitos.

Este procedimiento de análisis también se ha aplicado, dentro del método cuantitativo, con éxito al análisis de contenido de prensa (Navarro, Urchaga y otros, 2003; Polo, 2004; Quintana, 2004) (aquí las variables son categorías de análisis); así como con el método cualitativo, con material de obras narrativas y periodísticas de Manuel Vázquez Montalbán (tras un análisis con el programa informático Nudist, las variables son los categorías emergentes; Rey, 2005).

Lo importante de esta comunicación es comprobar el diálogo entre lo cuantitativo (como pueda ser el análisis de cluster) y lo cualitativo, ya que volvemos a las palabras escritas por los sujetos para interpretar los resultados. Se presenta un método que permite simultáneamente testar y construir teorías, ya que respeta el material cualitativo y logra realizar estadísticas multivariantes (que son las adecuadas para la realidad humana, la cual es compleja).

REFERENCIAS

- Berganza, M.R. y Ruiz, J.A. (coors.) (2005). *Investigar en comunicación*. Madrid: Mc Graw Hill
- Delgado, J. M. y Gutiérrez, J. (coord.) (1995). *Métodos y técnicas cualitativas de investigación en ciencias sociales*. Madrid: Síntesis.
- Navarro, C., Urchaga, J.D. y otros (2004). *La imagen de la discapacidad en la prensa de Castilla y León*. Salamanca: Junta de Castilla y León, Consejería de Familia e Igualdad de Oportunidades.
- Newman, I. y Benz, C.R. (1998). *Qualitative-Quantitative research methodology: exploring the Interactive continuum*. Carbondale: Illinois University Press.
- Deconchy, J.P. (1967) *Structure genetique de l'idee de Dieu chez des Catholiques Francais, garçons et filles de 8 à 16 ans*. Bruxelles: Lumen Vitae.
- Kerlinguer, F.N. y Howard, B.L. (2002) (4ªed, orig.1986) *Investigación del comportamiento. Métodos de investigación en ciencias sociales*. Madrid: Mc Graw-Hill.
- Polo Fernández, M^a. E. (2004) *La representación de los mayores en los periódicos de Castilla y León (1983-2001)*. Tesis doctoral inédita, Universidad Pontificia de Salamanca.
- Quintana Paz, N. (2004) *Políticas de comunicación y contraprestación editorial. Los grupos mediáticos en los gobiernos de la UCD y del PSOE*. Tesis doctoral inédita, Universidad Pontificia de Salamanca.
- Rey, P. (2005) *Análisis temático de la obra narrativa y periodística de Manuel Vázquez Montalbán*. Tesis doctoral inédita, Universidad Pontificia de Salamanca.
- Urchaga, J.D. (2005) *Evolución de la idea de Dios entre los 8 y 18 años*. Tesis doctoral inédita, Universidad Pontificia de Salamanca.
- Urchaga, J.D. (2007) Análisis de contenido. En Román Reyes (Dir.) *Diccionario Crítico de Ciencias Sociales*. Madrid y México: Ed. Plaza % Valdés.

Simplificación de los Análisis Morfométricos para Poblaciones de la Tortuga Dulceacuícola *kinosternon herrerae*

Gustavo Aguirre León¹, Erasmo Cázares Hernández², Blanca Hervier Martínez³

Universidad Veracruzana

¹gustavo.aguirre@inecol.edu.mx

²caeh72@hotmail.com

³bhervier@hotmail.com

Resumen

Introducción:

Los estudios morfométricos de tortugas son una herramienta para llevar a cabo tanto estudios descriptivos como de ecología de poblaciones. Sin embargo, generalmente no existe consenso en cuanto a las variables utilizables. En muchos casos el esfuerzo de muestreo es innecesariamente alto, tomándose mediciones redundantes, mientras que en otros casos ocurre lo contrario.

Kinosternon herrerae, es una tortuga endémica del este de México, y una de las especies menos estudiadas de la familia de los kinosternidos. Existen varios estudios que comprenden aspectos de descripción de las dimensiones corporales de esta tortuga (Shannon y Smith, 1949; Poglayen y Smith, 1958; Poglayen, 1965; Williams y Wilson, 1965; Pritchard, 1967, 1969). Recientemente se realizó un detallado estudio sobre la morfometría de una población de *Kinosternon herrerae* en un arroyo situado en el municipio de Xalapa, en el que se exponen medidas de 49 variables tomadas a partir de 80 individuos (Cázares-Hernández, 2007). Esta amplia colección de datos supone una oportunidad para discriminar cuáles son aquellas variables de las que se podría prescindir para reducir los costos de la toma de datos.

Objetivo:

El presente trabajo pretende contribuir a simplificar los estudios morfométricos de las poblaciones de *Kinosternon herrerae*, mediante la utilización de técnicas estadísticas que minimicen el número de variables a tomar en cuenta.

Metodología:

El estudio del cual partimos se llevó a cabo en un arroyo denominado “La Bomba”, que se encuentra en el municipio de Xalapa, Veracruz, al sureste de la capital del mismo nombre. Mediante un método de trampeo se capturaron 80 individuos, los cuales se liberaban en un corto periodo de tiempo, habiéndolos marcado antes para poder identificarlos en nuevas recapturas. Los individuos se diferenciaron en machos y hembras de acuerdo con una combinación de caracteres sexuales secundarios: el tamaño de la cola, la presencia de almohadillas de fricción en los miembros posteriores, el tamaño de la cabeza y la concavidad del plastron. Los individuos sin desarrollo conspicuo de estos caracteres fueron considerados inmaduros. Se registraron 49 variables morfométricas incluido el peso (tabla 1).

Tabla 1. Variables morfométricas de caparazón, plastron y cabeza de *Kinosternon herrerai* agrupadas según se realizaron los análisis.

| Grupo 1: CAPARAZÓN | | Grupo 2: ESCUDOS DEL PLASTRON | |
|------------------------------|-------------------------------------|-------------------------------|--|
| LC | Longitud del caparazón | LP | Longitud del plastron |
| AC ¾ | Ancho del caparazón, marginal ¾ | A GUL | Ancho Gular |
| AC 4 | Ancho del caparazón, marginal 4 | GUL | Escudo Gular |
| AC 5 | Ancho del caparazón, marginal 5 | HUM | Escudo Humeral |
| AC 6 | Ancho del caparazón, marginal 6 | PEC | Escudo Pectoral |
| AC 7 | Ancho del caparazón, marginal 7 | ABD | Escudo Abdominal |
| AC 8 | Ancho del caparazón, marginal 8 | FEM | Escudo Femoral |
| LV 1 | Longitud de escudo vertebral 1 | AN | Escudo Anal |
| LV 2 | Longitud de escudo vertebral 2 | AN 1 | Ancho de escudo Anal |
| LV 3 | Longitud de escudo vertebral 3 | AN 2 | Distancia en el extremo entre escudos anales |
| LV 4 | Longitud de escudo vertebral 4 | Grupo 3: FORMA DEL PLASTRON | |
| LV 5 | Longitud de escudo vertebral 5 | | |
| LCD 1 | Longitud de escudo costal derecho 1 | LPF | Longitud del Plastron Fisura |
| LCD 2 | Longitud de escudo costal derecho 2 | PUENTE D | Longitud del puente derecho |
| LCD 3 | Longitud de escudo costal derecho 3 | PUENTE I | Longitud del puente izquierdo |
| LCD 4 | Longitud de escudo costal derecho 4 | L LOB A | Longitud del Lóbulo Anterior |
| L NUC | Longitud del escudo nual | A LOB A | Ancho del Lóbulo Anterior |
| Grupo 4: ALTURAS | | L LOB P | Longitud del Lóbulo posterior |
| | | A LOB P | Ancho del Lóbulo posterior |
| ALT 1 | Altura anterior del caparazón | DIST 2 AN | Distancia 2 Anal |
| ALT 2 | Altura media del caparazón | L PUNTA AN | Longitud Punta Anal |
| ALT 3 | Altura posterior del caparazón | | |
| OTRAS VARIABLES NO AGRUPADAS | | | |
| A CAB | Ancho de la cabeza | AN 3 | Distancia entre extremo medio de escudos anales y borde del caparazón |
| L CAB | Longitud de la cabeza | AN 4 | Distancia ente el extremo del escudo anal derecho y el borde del caparazón |
| LC curv | Longitud curvilínea del caparazón | AN 5 | Distancia ente el extremo del escudo anal izquierdo y el borde del caparazón |
| AC curv | Ancho curvilíneo del caparazón | GROSOR AN | Grosor del Escudo Anal |
| CONC | Concavidad máxima del plastron | ANILLOS | Long de los anillos de crecimiento en el escudo abd. dch. |

Para simplificar los análisis de este estudio y reducir los costos posteriores en la toma de datos se llevaron a cabo una serie de análisis estadísticos a partir de los datos de animales adultos (27 machos y 46 hembras). Los promedios, desviaciones estándar y máximos y mínimos de estas variables se muestran en la tabla 2.

Tabla 2. Estadística descriptiva de las variables morfométricas de *Kinosternon herrerai* utilizadas.

| variable | N | media | mín. | máx. | DS | variable | N | media | mín. | máx. | DS |
|----------|----|--------|------|------|-------|----------|----|-------|------|------|------|
| LC | 68 | 130.99 | 108 | 165 | 11.88 | AC 4 | 68 | 82.86 | 70 | 108 | 7.01 |
| AC 3/4 | 68 | 73.64 | 61 | 100 | 6.80 | AC 5 | 68 | 88.44 | 76 | 110 | 6.59 |
| AC 6 | 68 | 88.78 | 75 | 109 | 6.70 | AC7 | 68 | 88.83 | 75 | 109 | 7.21 |
| VC | 68 | 86.82 | 73 | 103 | 7.52 | LV1a | 68 | 28.84 | 24 | 37 | 3.05 |
| ALT 1 | 65 | 45.03 | 35 | 60 | 5.37 | LV2a | 68 | 28.92 | 21 | 39 | 3.04 |
| ALT 2 | 65 | 47.31 | 39 | 59 | 4.37 | LV3a | 68 | 25.18 | 21 | 32 | 2.44 |
| ALT 3 | 65 | 42.50 | 33 | 54 | 4.15 | LV4a | 68 | 24.30 | 18 | 34 | 2.82 |
| LP | 68 | 109.51 | 86 | 134 | 8.37 | LV5a | 68 | 27.00 | 17 | 36 | 4.06 |
| LPF | 68 | 103.79 | 87 | 130 | 7.66 | LCD1a | 68 | 39.69 | 20 | 51 | 4.90 |
| GUL | 68 | 15.73 | 5 | 20 | 2.48 | LCD2a | 68 | 29.75 | 24 | 56 | 4.53 |

| variable | N | media | mín. | máx. | DS | variable | N | media | mín. | máx. | DS |
|----------|----|--------|------|------|-------|-----------|----|-------|------|------|------|
| HUM | 68 | 15.59 | 11 | 21 | 2.02 | LCD3a | 68 | 26.20 | 18 | 33 | 3.08 |
| PEC | 68 | 8.16 | 2 | 19 | 2.99 | LCD4a | 68 | 27.79 | 18 | 36 | 3.87 |
| ABD | 68 | 28.54 | 21 | 37 | 3.36 | LONG NUC | 66 | 5.70 | 2 | 10 | 1.30 |
| FEM | 68 | 12.70 | 8 | 20 | 2.38 | A GUL | 68 | 19.40 | 11 | 28 | 3.10 |
| AN | 68 | 27.47 | 19 | 34 | 2.96 | DIST 2 AN | 68 | 33.30 | 27 | 41 | 2.77 |
| GROS A | 63 | 2.13 | 1 | 3 | 0.57 | LONG P AN | 68 | 10.49 | 6 | 16 | 2.51 |
| PUENT D | 68 | 21.10 | 9 | 30 | 2.73 | AN 1 | 68 | 42.36 | 35 | 55 | 4.18 |
| PUENT I | 68 | 21.24 | 9 | 28 | 2.77 | AN 2 | 68 | 17.24 | 4 | 34 | 4.89 |
| PESO | 66 | 355.98 | 215 | 780 | 95.83 | AN 3 | 67 | 23.90 | 16 | 40 | 5.40 |
| LOBANT L | 64 | 37.47 | 28 | 45 | 3.46 | AN 4 | 67 | 18.79 | 12 | 34 | 4.55 |
| LOBANT A | 64 | 53.98 | 47 | 69 | 4.20 | AN 5 | 67 | 18.64 | 13 | 34 | 4.67 |
| LOBP L | 64 | 47.85 | 37 | 58 | 3.67 | CONC | 68 | 0.89 | 0 | 6 | 1.36 |
| LOBP A | 64 | 47.84 | 39 | 60 | 3.77 | CAB A | 48 | 32.86 | 26 | 43 | 4.29 |
| LC curv | 67 | 149.66 | 123 | 188 | 13.90 | CAB L | 48 | 46.69 | 34 | 61 | 6.91 |
| AC curv | 67 | 127.19 | 107 | 155 | 10.27 | | | | | | |

Para poder identificar aquellas variables que más importancia tengan a la hora de describir la variabilidad de cada uno de los aspectos morfológicos de la tortuga (parte superior, perfil, parte inferior, etc.) las variables se agruparon como se indica en la tabla 1. Esto permite obtener resultados independientes de cada parte del animal, de manera que posteriormente sea más valiosa la comparación de unas poblaciones con otras de la misma especie o especies diferentes.

Para los análisis se usaron los Software Minitab13 y Statistica 7.0. Un análisis de componentes principales (ACP) por grupos de variables nos permitió reducir el número de variables usando toda la información, seleccionando aquellas que en conjunto contribuían al 90% de la formación de la primera componente principal. Con un análisis discriminante de clasificación cuadrática (ADC) detectamos la probabilidad de cometer un error al sexar un nuevo individuo. Para este análisis usamos las variables seleccionadas con el ACP y las variables no agrupadas. De las variables no agrupadas se seleccionaron las que tenían mayor importancia en la discriminación de sexos, además de las imprescindibles para cualquier estudio morfométrico (peso y anillos de crecimiento). De esta manera redujimos el número de variables a tomar en cuenta, minimizando la pérdida de información.

Resultados:

A partir del ACP obtenemos el primer componente principal para cada grupo de variables. Para el grupo 1 el primer componente principal explica un 85% de la variabilidad de los datos, en el 2 explica el 63% en el 3 explica el 74% y para el 4 el 85%. Las variables seleccionadas, la proporción en la que contribuyen a la variabilidad total del primer componente, y su contribución acumulada son:

| GRUPO 1 | Aportación al comp. 1 | Aportación acumulada |
|---------|-----------------------|----------------------|
| LC | 0.29 | 0.29 |
| VC | 0.11 | 0.40 |
| AC 4 | 0.10 | 0.50 |
| AC 7 | 0.10 | 0.60 |
| AC 3/4 | 0.09 | 0.69 |
| AC 6 | 0.09 | 0.78 |
| AC5 | 0.09 | 0.87 |
| LV 5 | 0.03 | 0.90 |

| GRUPO 2 | Aportación al comp. 1 | Aportación acumulada |
|---------|-----------------------|----------------------|
| LP | 0.63 | 0.63 |
| AN 1 | 0.13 | 0.76 |
| AN 2 | 0.09 | 0.85 |
| ABD | 0.06 | 0.91 |

| GRUPO 3 | Aportación al comp. 1 | Aportación acumulada |
|-----------|-----------------------|----------------------|
| LPF | 0.54 | 0.54 |
| LOB ANT A | 0.12 | 0.66 |
| LOB P A | 0.10 | 0.76 |
| LOB P L | 0.08 | 0.84 |
| LOB ANT L | 0.06 | 0.90 |

| GRUPO 4 | Aportación al comp. 1 | Aportación acumulada |
|---------|-----------------------|----------------------|
| ALT 1 | 0.48 | 0.48 |
| ALT 2 | 0.28 | 0.76 |
| ALT 3 | 0.24 | 1.00 |

Partiendo de estas variables y de las variables no agrupadas, un análisis de clasificación cuadrática (ADC) nos clasifica a los organismos correctamente con una confiabilidad de 0.98 (0.97 para machos y 1.00 para hembras). Las variables que mejor clasifican a los organismos por sexo son la longitud del lóbulo anterior y el ancho del lóbulo posterior de plastrón, el ancho de la cabeza, la concavidad y el ancho curvilíneo del caparazón. Estas tres últimas medidas son seleccionadas entre las variables que quedaban sin agrupar, por determinar en gran medida el dimorfismo sexual. Además se deben tener en cuenta el peso y la longitud de los anillos de crecimiento por tener mucha importancia para algunas aplicaciones de los estudios morfométricos de quelonios.

Conclusiones:

El empleo de un gran número de variables morfométricas es necesario a la hora de describir una especie que no ha sido descrita morfométricamente con anterioridad. Sin embargo, una vez realizado un estudio morfométrico preliminar se pueden discriminar aquellas variables de las que se puede prescindir en estudios posteriores, en este caso de la especie *Kinosternon herrerai*, con el fin de reducir los costos en la toma de datos. Un análisis de componentes principales por grupos de variables y un análisis discriminante nos permitieron reducir de 49 a 25 las variables a considerar en estudios posteriores.

De las variables correspondientes a la parte superior del caparazón únicamente se recomienda medir el quinto escudo vertebral. El resto de escudos del caparazón no aportan realmente mucha información a la hora de explicar la variabilidad de los individuos, como arroja el análisis de componentes principales. Las variables que sí son importantes medir en el caparazón son por supuesto su longitud, así como el ancho. El ancho del caparazón medido a seis niveles diferentes nos da una idea de la forma del caparazón de la tortuga, sin

embargo todas estas medidas están muy correlacionadas, y se puede considerar reducirlas a tres. Se recomienda medir tres alturas de caparazón, así como el ancho curvilíneo. Estas medidas, junto a la longitud del caparazón, determinan el perfil de la tortuga, muy diferente en machos y hembras adultos.

La forma y el tamaño del plastrón también son diferentes entre sexos por lo que se deben medir su longitud total y la longitud y el ancho de los lóbulos anterior y posterior. Por otro lado es importante medir el plastrón a nivel de la fisura, puesto que la profundidad de la fisura es una característica que también puede variar bastante entre machos y hembras debido a la necesidad de espacio para la ovipostura. Por lo mismo se medirá el ancho de la placa anal a dos niveles (AN 1 y AN 2, ver tabla 1). De los escudos del plastron, no es necesario medir más que el escudo abdominal a nivel de la fisura central, que además será el empleado para medición de anillos de crecimiento.

Del resto de medidas, el ancho de cabeza es también un elemento de dimorfismo, que está relacionada con el tamaño de presa. Y por otro lado es importante tener en cuenta la concavidad del plastrón, mayor en machos de manera que se facilite la cópula. Finalmente, en cualquier estudio morfométrico se deben considerar las medidas del peso, así como de los anillos de crecimiento.

Estudios de este tipo son útiles por un lado para minimizar los esfuerzos en campo y facilitar la toma de datos, disminuyendo la probabilidad de error. Por otro lado se necesitan homogeneizar los trabajos de morfometría en quelonios, de manera que resulten comparables unos con otros.

Referencias:

- Cázares Hernández, E. (2007). Aspectos ecológicos de la tortuga *Kinosternon herrerae*, Stejneger, 1925 (Reptilia: Testudines: Kinosternidae) en el arroyo "La Bomba", municipio de Xalapa, Veracruz, México. Tesis de Licenciatura. Universidad Veracruzana. Xalapa.
- Poglayen, I. (1965). Observations on Herrera's mud turtle, *Kinosternon herrerae* Stejneger. *International Zoology Yearbook* 5, 171-173.
- Poglayen, I., Smith H. M. (1958). Noteworthy reptiles from México. *Herpetologica* 14, 11-15.
- Pritchard, P. C. H. (1967). Living turtles of the world. T.F.H. Publ. Inc., Jersey city, New Jersey. 288 pp.
- Pritchard, P. C. H. (1969). Herrera's mud turtle. *International Turtle and Tortoise Society Journal* 3, 6-9, 35.
- Shannon, F. A., Smith H. M. (1949). Herpetological results of the University of Illinois field expedition, spring, 1949. I. Introduction. Testudines, Serpentes. *Transactions of the Kansas Academy of Science* 52, 499-514.
- Williams, K. L., Wilson L. D. (1965). Noteworthy Mexican reptiles in the Louisiana State University Museum of Zoology. *The Proceedings of the Louisiana Academy of Science* 28, 127-130.

Estudio Multivariante de las Características del Turista Español que Visita México

Jaime Egado¹, Purificación Galindo²

¹ Departamento de Estadística. Universidad de Salamanca. jegido@usal.es

² Departamento de Estadística. Universidad de Salamanca. pgalindo@usal.es

1. Introducción

El turismo contribuye en un gran porcentaje en el producto interior bruto (PIB) de algunos países. Para México, décimo país del mundo receptor de turismo [4] con 22,6 millones de turistas internacionales en el año 2008 de los cuales 11 millones han llegado por vía aérea, representa más del 8,2% de su PIB en 2006 [5], generando dos millones de empleos. El turismo se ha convertido en uno de los impulsores del desarrollo económico y social del país.

España, segundo receptor de turismo mundial [4], es el sexto país que más turistas aporta a México, 371 mil viajeros en el año 2008.

Nuestro análisis estudia las características de los residentes españoles que viajan México.

2. Materiales y métodos

Los datos analizados fueron obtenidos en el Instituto de Estudios Turísticos (IET) de España, que es la entidad responsable de generar los datos desde el lado de la demanda. Entre las encuestas que realiza el IET está la encuesta de Movimientos Turísticos en Fronteras (FRONTUR) cuyo objetivo es cuantificar mes a mes el número de entradas a España por las distintas vías de acceso. Se realizan 360.000 encuestas autocumplimentadas al año en los aeropuertos españoles [2].

Las características estudiadas y sus niveles son:

- Tipo de alojamiento: complejo turístico con servicio de hotel, hoteles y similares, vivienda gratuita (propia, familiares), vivienda alquilada, otros alojamientos.
- Edad: menor de 15 años, de 15 a 24, de 25 a 44, de 45 a 64, mayor de 64 años.
- Motivo del viaje: ocio y vacaciones, trabajo o negocios ferias y congresos, personal (familiares, salud, compras), otros motivos.
- Paquete turístico: sin paquete, con paquete.
- Duración de la estancia: ninguna noche, 1 noche, de 2 a 3, de 4 a 7, de 8 a 15, más de 15 noches.
- Sexo: hombre, mujer.
- Tipo de vuelo: chárter, regular

Para el análisis de las características se ha utilizado el número de viajeros, pernoctaciones, estancia media, variación interanual de viajeros y pernoctaciones.

Realizamos un análisis multivariante HJ-Biplot (Galindo, 1986), estableciendo las características de los viajeros y analizando luego las variaciones de cada mes.

El análisis HJ-Biplot es una representación gráfica multivariante de datos. Nos permite recoger la estructura de covariación entre variables y representar en un espacio de baja dimensión, en el mismo sistema de referencia, las variables observables y las características de los viajeros, con máxima calidad de representación.

Los gráficos Biplot contienen puntos (características de los viajeros) y vectores (variables medidas). En primer lugar se evalúa la absorción de inercia recogida con los ejes factoriales. Si la absorción de inercia es suficientemente alta se analizan los planos factoriales, en los que sólo son interpretables los puntos con alta calidad de representación.

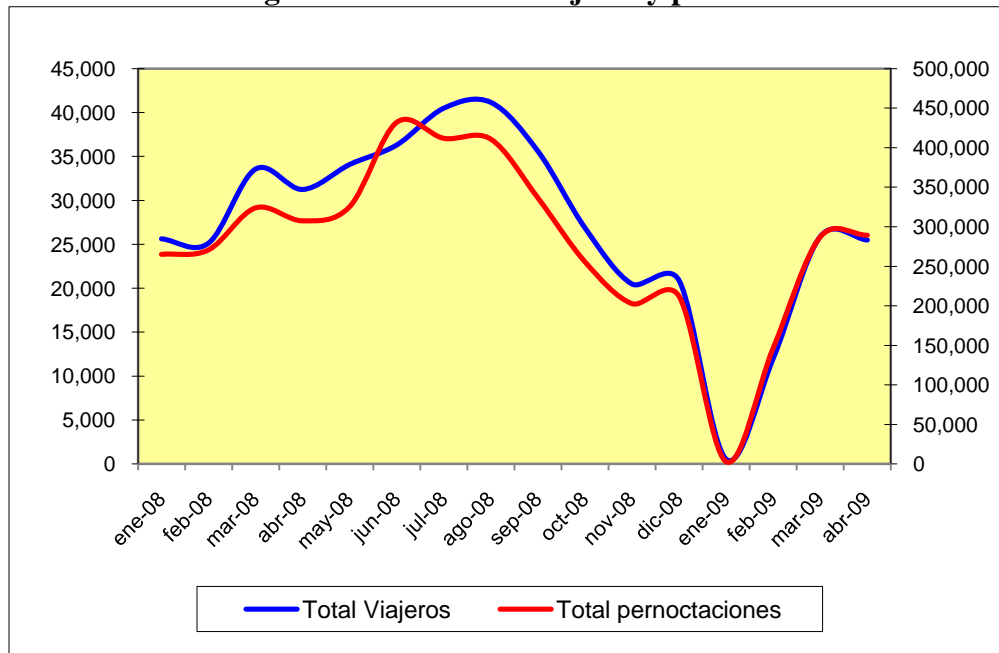
La interpretación sobre el plano se basa en: a) la proximidad de los puntos indica similitud entre ellos. b) la longitud del vector indica la variabilidad aportada por la variable. c) el ángulo entre dos vectores indica la relación entre ambas variables; un ángulo pequeño indica fuerte relación y un ángulo recto indica independencia de las dos variables. d) cada punto se puede proyectar sobre cada vector; si están en la dirección de la punta del vector y separados del origen indica valores altos sobre dicha variable. e) los ejes factoriales representan gradientes y se interpretan evaluando las contribuciones de cada variable al gradiente latente.

El software utilizado es el programa Multbiplot (Villardón, 2007).

3. Análisis

En la Figura 2 podemos ver la evolución que han sufrido el número de viajeros de residentes españoles y las pernoctaciones realizadas por ellos en México desde enero de 2008, según los datos facilitados por el IET. Se observa como las magnitudes crecen hasta el verano, momento en que comienza una fuerte caída hasta el mes de enero de 2009. También se observa una disminución más acusada en el número de viajeros que en las pernoctaciones.

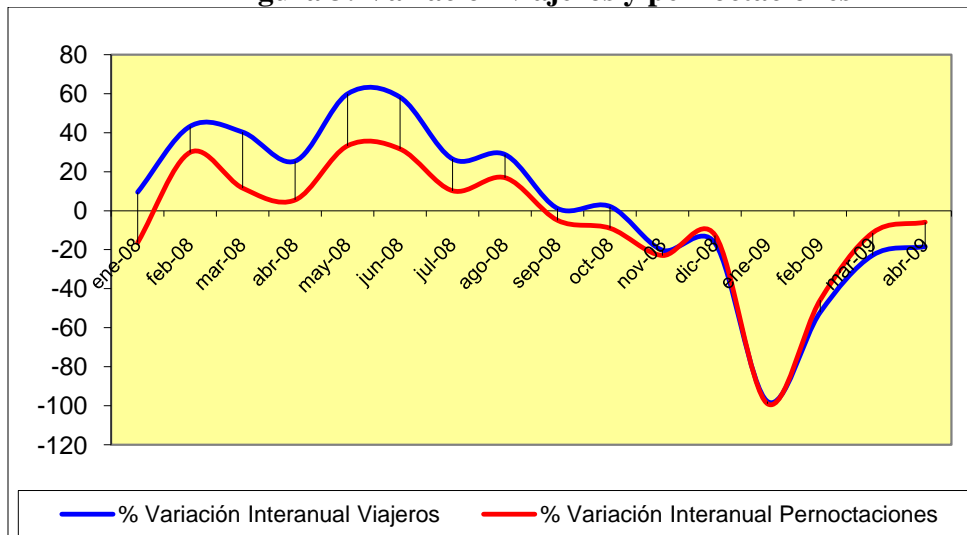
Figura 2: Número de viajeros y pernoctaciones



Los datos existentes hasta abril de 2009 todavía no recogen los efectos producidos por la aparición de la gripe A.

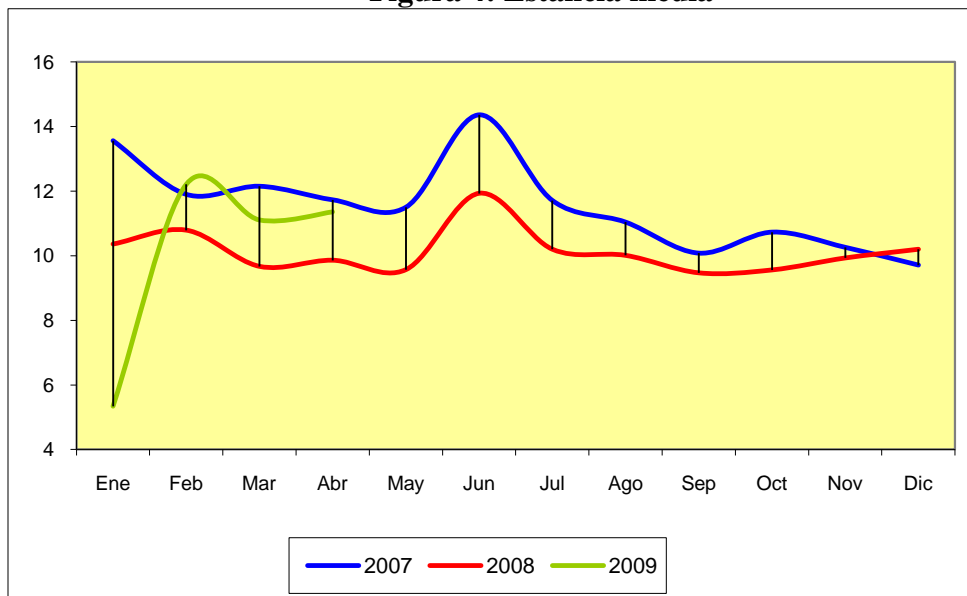
La Figura 3 ratifica lo anterior y que a partir del verano los viajeros tienen mayor caída que las pernoctaciones.

Figura 3: Variación viajeros y pernoctaciones



La estancia media ha caído durante el año 2008 pero se ha recuperado durante el año 2009, salvo enero que tuvo una caída muy fuerte. Figura 4.

Figura 4: Estancia media



Para analizar las características de los viajeros utilizamos un análisis multivariante HJ-Biplot, con los datos de los años 2007 y 2008 de forma conjunta. Con una explicación del 99% de absorción de inercia, como se puede apreciar en la Figura 5, el número de turistas y pernoctaciones viene definido por las personas que viajan en vuelo regular, por motivos de ocio, que contratan paquete turístico, en hotel y tienen entre 25 y 44 años. La estancia media mayor de 15 noches se realiza en vivienda alquilada y vivienda gratuita, propia o de amigos.

De forma similar, y en sentido contrario, el menor número de viajeros y pernoctaciones vienen caracterizados por estancias cortas, alojamientos en complejos turísticos, por motivos de trabajo, personales y otros, en vuelos chárter y edad mayor de 45 años y menor de 25.

Figura 5: Características de los turistas

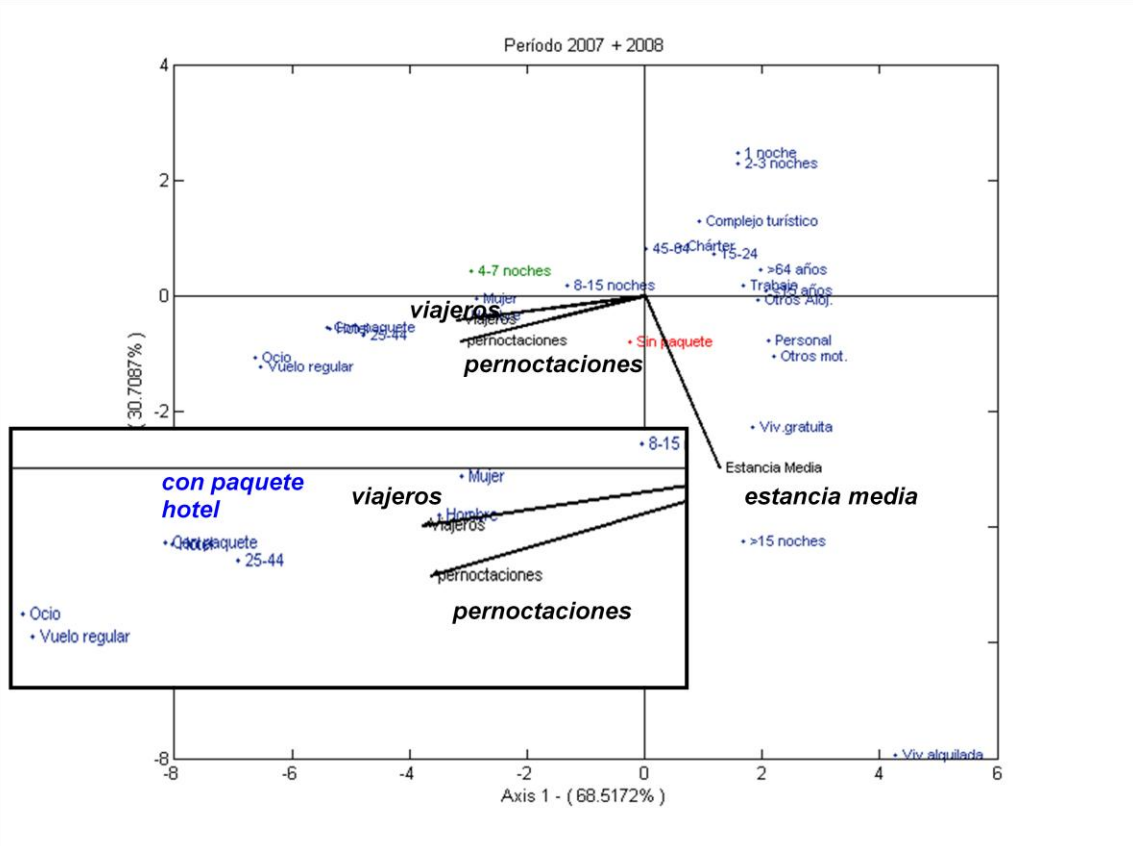
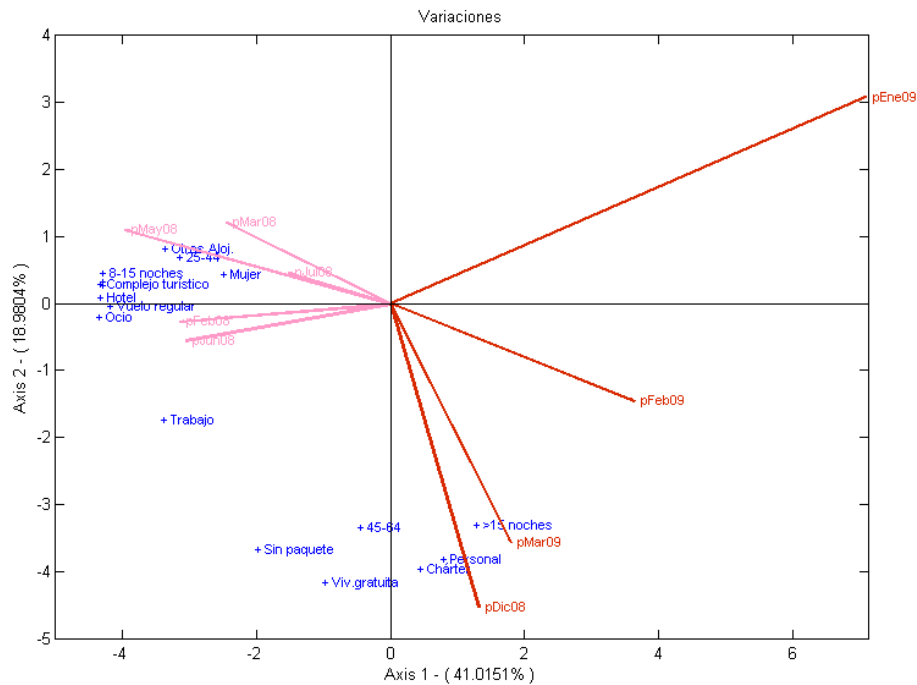


Figura 6: Variación en las pernoctaciones



Las variaciones en el número de viajeros y pernoctaciones se manifiestan de forma similar. En la Figura 6 se observan que las variaciones positivas en las pernoctaciones están caracterizadas por estancias en hotel, por motivo de ocio, en vuelo regular, entre 8 y 15 noches, de 25 a 44 años, en los meses de febrero, marzo, mayo, junio y julio de 2008, lo que concuerda con lo visto en la Figura 3.

Los descensos de diciembre de 2008 y marzo de 2009 se deben a viajes sin paquete turístico, en vuelo chárter, por motivos personales, a vivienda gratuita, por más de 15 noches y edad de 45 a 64 años. Los descensos de febrero y marzo de 2009 se deben a disminuciones en todos los conceptos.

4. Conclusiones

El turista tipo viaja con paquete turístico, por motivos de ocio, vuelo regular, hotel, con una duración de estancia entre 4 y 15 días, y edad entre 25 y 44 años.

La mayor estancia media es en vivienda alquilada y vivienda gratuita, sin paquete turístico, durante más de 15 noches y motivos personales.

Existe un incremento de viajeros y pernoctaciones hasta agosto de 2008 produciéndose un acusado descenso con mínimo en enero de 2009. En diciembre la variación es más significativa en estancias mayores de 15 días, en personas entre 45 y 64 años, sin paquete turístico, en vuelo chárter, vivienda gratuita y por motivos personales.

La variación positiva en la estancia media es desde septiembre de 2008 hasta marzo de 2009, excepto enero de 2009, en viviendas alquiladas, menores de 15 años y durante más de 15 noches.

La variación negativa para el resto de los meses se debe al descenso en hoteles y complejos turísticos, motivos de ocio y trabajo, menos de 15 noches, con paquete, tanto en vuelo regular como chárter, y de edades entre 25 y 44 años, es decir, en todas las características del turismo.

REFERENCIAS

- [1] Galindo, M.P. (1986) *Una alternativa de representación simultánea: HJ-biplot*. *Questío*, 10(1), 13-23.
- [2] Instituto de Estudios Turísticos (IET), *Ficha técnica de la Estadística de Movimientos Turísticos en Fronteras (FRONTUR)*. www.iet.tourspain.es.
- [3] Instituto de Estudios Turísticos (IET), *Movimientos Turísticos en Fronteras (FRONTUR). Informe anual 2007*. www.iet.tourspain.es.
- [4] Organización Mundial del Turismo (OMT). *Panorama del turismo internacional. Edición 2008*.
- [5] Secretaría de Turismo de México. *Análisis del turismo*. Diciembre 2008. datatur.sectur.gob.mx.
- [6] Villardón, J.L. (2007) *Programa MULTBILOT v.alpha 3.1*. Departamento de Estadística. Universidad de Salamanca.

Regresión Logística Ordinal: Aplicación en Niveles de Estrés

Luis F. Grajales¹, Iván Mendivelso²

¹Departamento de Estadística. Universidad Nacional de Colombia. lfgrajalesh@unal.edu.co

²Universidad Nacional de Colombia. vania.unal@gmail.com

Agosto 2009, Veracruz, México, 2do. Encuentro Iberoamericano de Biometría

Índice

| | |
|---|----------|
| 1. Datos, Variable Respuesta y covariables | 1 |
| 1.1. Distribución multinomial | 2 |
| 1.2. El modelo <i>logit</i> acumulativo | 2 |
| 1.3. Regresión logística ordinal | 3 |
| 1.4. Verificación de supuestos: examinando las sumas acumuladas | 3 |
| 1.5. Estadístico razón de verosimilitud | 3 |
| 1.6. Depuración y selección de variables | 4 |
| 2. Bondad de ajuste | 4 |
| 2.1. Dos medidas de bondad de ajuste | 4 |
| 2.2. Nivel de asociación del modelo | 4 |
| 3. Resultados | 4 |
| 3.1. Covariables finales en el modelo | 4 |
| 3.2. Un ejemplo de cálculo de valores esperados | 5 |
| 3.3. Odds Ratios | 5 |

1. Datos, Variable Respuesta y covariables

Se encuestaron 1300 personas de 18 diferentes ciudades colombianas; a estas personas se les preguntó por su nivel de estrés y por otras preguntas concernientes a diferentes aspectos de su vida, tales como actitudes, opiniones, intereses, etc.

El nivel de estrés de las personas fue abordada como una variable latente [4]. La variable está medida mediante la siguiente escala y se formuló su pregunta así: En una escala de "1" a "5", donde 1 es poco estresado y 5 muy estresado, ¿qué tan estresado se encuentra usted en estos días? (ÚNICA RESPUESTA)

POCO ESTRESADO (1 2 3 4 5) MUY ESTRESADO

Mediante este estudio se ha buscado el acercamiento al conocimiento e identificación de ciertos perfiles de habitantes de algunas ciudades de Colombia. Para este efecto, se miden variables en módulos, que buscan captar las actitudes, opiniones e intereses, preguntando por costumbres, tradiciones, comportamientos financieros, gasto de tiempo de ocio, uso de tecnología, entre otros, con el fin de perfilar y caracterizar el *target* antedicho. Dichos módulos buscan abarcar, entre otros aspectos:

- Salud (entidades médicas, prevención, etc.)
- Comida (nutrición, tendencia light, etc.)
- Aspectos sociales
- Finanzas (productos financieros, ahorro, etc.)
- Tecnología (Internet, computadores, contacto con aparatos, etc.)

1.1. Distribución multinomial

La regresión ordinal saca provecho de la existencia de una relación de orden entre las categorías de la variable respuesta Y .

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}, \text{ con } \sum_{j=1}^J y_j = n.$$

Para este vector de variables, la distribución multinomial está dada por:

$$f(\mathbf{y}|n) = \frac{n!}{y_1! y_2! \dots y_J!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_J^{y_J},$$

denotada por $M(n, \pi_1, \pi_2, \dots, \pi_J)$, y con $\sum_{j=1}^J \pi_j = 1$; $E(Y_j) = n\pi_j$; $\text{var}(Y_j) = n\pi_j(1 - \pi_j)$ y $\text{cov}(Y_i, Y_j) = -n\pi_i\pi_j$.

La distribución multinomial no pertenece a la familia exponencial uniparamétrica. No obstante, la relación que guarda la distribución multinomial con la distribución Poisson y, bajo una distribución condicional, se puede tener una expresión perteneciente a la familia exponencial.

1.2. El modelo *logit* acumulativo

Modelo *logit* acumulativo:

$$\frac{\pi_1}{\pi_2}, \frac{\pi_2}{\pi_3}, \dots, \frac{\pi_{J-1}}{\pi_J}.$$

El modelo logit de categorías adyacentes es:

$$\log\left(\frac{\pi_j}{\pi_{j+1}}\right) = X_j^T \beta_j = \beta_{0j} + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}$$

En este caso, $j = 1, 2, 3, 4, 5$, ya que son 5 diferentes niveles de estrés que se medirán y los π_j hacen referencia a la probabilidad de que un individuo caiga en la categoría de respuesta j .

1.3. Regresión logística ordinal

$\exp(\beta_j + X^T \beta) = P(\text{una persona puntúe } j \text{ o menos}) / P(\text{una persona puntúe mayor que } j)$

Obsérvese que existen $J - 1 = 4$ cocientes que involucran la variable respuesta, existiendo un intercepto asociado directamente a cada una de las categorías, con excepción de la quinta. Un modelo para un conjunto de p variables (una dependiente y $p - 1$ variables explicativas), esatría definido por:

$$\log \frac{\pi_1 + \pi_2 + \dots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \dots + \pi_J} = \beta_{0j} + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}.$$

Se tiene que cada uno de los logit para cada categoría $j = 1, \dots, J$ y p variables Y, X_1, \dots, X_{p-1} , tiene su propio intercepto β_{0j} ; sin embargo, el resto de betas $\beta_1, \beta_2, \dots, \beta_{p-1}$ son los mismos para todas las ecuaciones. Esto se traduce en que los efectos de las variables respuesta son los mismos para las $J - 1$ diferentes funciones logit.

1.4. Verificación de supuestos: examinando las sumas acumuladas

El porcentaje acumulado presenta *paralelismos* y sólo se interceptan al final, cuando ambos porcentajes acumulan 100%. Se aplicará el respectivo test chi-cuadrado para determinar la presencia de paralelismo entre las covariables explicativas.

1.5. Estadístico razón de verosimilitud

La función de verosimilitud es la probabilidad de la muestra, vista como una función del parámetro una vez los datos ya han sido observados. La prueba de razón de verosimilitud determina los valores del parámetro que maximizan la función de verosimilitud, bajo el supuesto de que H_0 es verdad, sobre la condición más general de que H_0 pueda o no ser cierta (o más comúnmente, *no rechazar*) [2]:

$-2Ln(\text{máxima verosimilitud bajo } H_0) / (\text{máxima verosimilitud sin restricciones a los parámetros})$

1.6. Depuración y selección de variables

Para este trabajo se compararon dos modelos propuestos por una psicóloga experta en el tema de estrés. A continuación se tienen los resultados del mejor modelo.

2. Bondad de ajuste

2.1. Dos medidas de bondad de ajuste

Se realizaron las pruebas de bondad de ajuste *Pearson* y *Deviance* con base en los valores observados y los esperados:

$$\chi^2 = \sum \sum_y \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$D = 2 \sum \sum O_{ij} (\ln(O_{ij}) - \ln(E_{ij}))$$

El modelo fue seleccionado, entre otras características de buen desempeño, por presentar p-valor para el *Deviance* test de 0.482 y, para *Pearson*, de 0.446. Dichos valores dan luz verde al siguiente paso de interés para continuar con la evaluación del modelo.

2.2. Nivel de asociación del modelo

Para estos efectos, Cox y Snell, Nagelkerke y McFadden, proponen 3 pseudo R-cuadrados, para evaluar la adecuación del modelo elegido. Estos se interpretan como el porcentaje de explicación de la respuesta al tener este modelo, comparado con el modelo sin covariables. [4]:

En el caso del modelo seleccionado, los pseudo- R^2 resultaron así:

- $R_{CS}^2 = 0,039$
- $R_N^2 = 0,041$
- $R_M^2 = 0,013$

3. Resultados

3.1. Covariables finales en el modelo

Al mirar la Tabla 1, se enuncia el modelo ajustado, como ejemplo, en la primera relación:

$$Ln \left[\frac{\pi_1}{\sum_{i=1}^4 \pi_i} \right] = -1,14 + 0,10X_1 + 0,71X_{21} + 0,62X_{22} + 0,55X_{23} + 0,34X_{32} - 0,36X_{41}$$

con:

- X_1 := número de menores de edad pertenecientes al hogar del encuestado.
- X_{21} := necesidad fuerte de reducir deudas.
- X_{22} := necesidad moderada de reducir deudas.
- X_{23} := necesidad leve de reducir deudas.
- X_{32} := madre.
- X_{41} := actitud preocupada por su alimentación cuando come por fuera de casa.

3.2. Un ejemplo de cálculo de valores esperados

$$P(Y \leq 1) = \frac{1}{1 + e^{1,137}} = 0,243$$

3.3. Odds Ratios

Tabla 1.

| Variables | β_i | p-value | Lim. inf. β_i | Lim. sup. β_i | $\frac{1}{1+exp(\beta_{01})}$ | Odds * β_{02} | Odds * $\beta_{02} + \beta_1 X$ | OR |
|------------|-----------|---------|---------------------|---------------------|-------------------------------|---------------------|---------------------------------|-------|
| $Y \leq 1$ | -1,137 | 0,000 | -1,581 | -0,714 | 0,243 | | | |
| $Y \leq 2$ | 0,216 | 0,316 | -0,198 | 0,621 | 0,554 | | | |
| $Y \leq 3$ | 1,217 | 0,000 | 1,398 | 2,216 | 0,860 | | | |
| $Y \leq 4$ | 3,000 | 0,000 | 2,825 | 3,354 | 0,958 | | | |
| X_1 | 0,097 | 0,028 | -1,281 | 0,122 | 0,860 | 8,152 | 5,584 | 1,102 |
| X_{21} | 0,712 | 0,000 | 0,316 | 1,108 | 0,860 | 8,152 | 3,018 | 5,038 |
| X_{22} | 0,623 | 0,003 | 0,212 | 1,394 | 0,860 | 8,152 | 3,320 | 1,864 |
| X_{23} | 0,555 | 0,020 | 0,087 | 1,023 | 0,860 | 8,152 | 3,565 | 1,725 |
| X_{32} | 0,343 | 0,013 | 0,091 | 0,595 | 0,860 | 8,152 | 4,388 | 1,422 |
| X_{41} | -0,360 | 0,002 | -0,286 | -0,132 | 0,860 | 8,152 | 5,305 | 0,208 |

Bibliografía

- [1] Dobson, A. J. *An Introduction to Generalized Linear Models* Chapman & Hall/CRC (2003, 2008).
- [2] Agresti, Alan *An Introduction to Categorical Data Analysis* Wiley series in probability and statistics (2007), 1-50.
- [3] Hosmer y Lemeshow *Applied Logistic Regression* New York: Wiley (2000).
- [4] Campbell, J. Quick, J., Nelson, D., Hurrell, J. Jr. *Preocupante Stress Management in Organizations* American Psychological Association (1997).

Descripción estadística de la evaluación clínica y neurofisiológica del Síndrome de Guillain-Barré en México

Juan José Méndez Castillo¹, Luis Angel Ruano Calderón², Mariangeles Moreno Barjollo³, Lucas López Segovia^{3,4}

¹Hospital de Alta Especialidad “Dr. Juan Graham Casasús”. Villahermosa, Tab. México. neu8ro-laboratorio@yahoo.com

²Jefatura de Enseñanza. Hospital General. Durango, México. laruanoc@hotmail.com

³Universidad Politécnica de Cataluña. Barcelona, Esp. maangeles_moreno@yahoo.com

⁴División Académica de Ciencias Básicas. Universidad Juárez Autónoma de Tabasco. Villahermosa Tab., México. llopez@ujat.mx

Resumen

En esta comunicación se presenta una descripción estadística de la evaluación clínica y neurofisiológica del síndrome de Guillen-Barré sobre pacientes ingresados en el Instituto Nacional de neurología y Neurocirugía de México durante los años 1993 a 2002. Los resultados preliminares indican que la variantes del Síndrome de Guillan Barre en México son diferentes a lo reportado por la literatura internacional tradicional de los países occidentales y guardan más similitud a lo reportado en comunidades de China.

1. Introducción

El SGB es una de las principales neuropatías desmielinizantes predominantemente motoras; representa el 25% al 40% de todas las polineuropatías de la edad adulta y la mitad de los pacientes refieren el antecedente de una infección de vías respiratorias altas en los días previos al inicio del cuadro. Se caracteriza por debilidad muscular que puede ser generalizada o bien iniciar en las extremidades inferiores con progresión distal a proximal de las mismas y después afecta otros territorios. Tras un corto período de estabilidad se inicia la recuperación, que suele ser completa a los 3-6 meses. La mortalidad en los casos más graves es del 5%. El 10% de los pacientes queda con secuelas permanentes, esto depende de la edad y de la lesión axonal producida. Desde el punto de vista clínico, en la actualidad se definen diversas variantes del SGB, cada una con sus criterios diagnósticos específicos: AIDP (polineuropatía desmielinizante inflamatoria aguda), AMDN (neuropatía desmielinizante motora aguda), AMAN (neuropatía axonal motora aguda), AMSAN (neuropatía axonal sensitivo-motor aguda) y el síndrome de Miller-Fisher. La realización de estudios neurofisiológicos son un complemento indispensable para el diagnóstico de SGB. La primera anomalía suele ser la prolongación o ausencia de las ondas F, enlentecimiento, incluso con bloqueo de la velocidad de conducción y se encuentran aumentadas las latencias dístales, con amplitudes disminuidas, degeneración axonal secundaria y potenciales de denervación en el estudio de electromiografía, hallazgo de valor pronóstico ya que indica la existencia de lesión axonal secundaria, pero provee una muy limitada evidencia de desmielinización, que es la principal característica del SGB. En los países occidentales se han descrito un patrón con forma motora pura con una tendencia a debilidad distal, sin afectar a los nervios craneales. Este patrón se ha ligado al incremento de infección por *Campylobacter jejuni* y a la presencia de anticuerpos anti-GM1. En China se ha descrito un patrón motor puro; la EMG en estos pacientes mostró consistentemente un

patrón motor puro y se encontró en niños y adultos jóvenes. Los reportes de China describen pacientes con parálisis flácida aguda con características que recuerdan a los pacientes con SGB desmielinizante, pero las manifestaciones epidemiológicas difieren de los pacientes vistos en América del Norte y Europa.

2. Descripción del estudio

Objetivo

Con el objetivo de caracterizar clínica y electrofisiológicamente a los pacientes con SGB y determinar la variante más frecuente en nuestro medio, se realizó un estudio de pacientes con SGB ingresados en el Instituto Nacional de neurología y Neurocirugía de México (INNN), siendo este Hospital referencia en la ciudad de México. Los pacientes incluidos en el estudio fueron todos los pacientes ingresados al INNN con diagnóstico de SGB, en su etapa aguda, con expediente clínico y resultados de estudios electrofisiológicos completos de 1993 a 2002. Se eliminaron las formas secuelas de SGB y otras polineuropatías.

Variables de estudio

Se tomaron en cuenta variables epidemiológicas y clínicas además de los resultados de estudios paraclínicos. De los estudios electrofisiológicos se tomaron en cuenta los resultados de los nervios motores mediano, cubital, tibial y peroneo motores y para los nervios sensitivos se evaluaron mediano, cubital y sural, tabulando las latencias distal y proximal, amplitudes, áreas bajo la curva, así mismo se evaluaron las Velocidad de Conducción Nerviosa (VNC) motora y sensitiva, las alteraciones y los hallazgos en la EMG con aguja.

Método

Los estudios fueron realizados previamente con un computador con programa de neuroconducción motora y sensitiva marca Nicolet, modelo Viking IV, para determinar la VCN motora y sensitiva en los nervios periféricos descritos y las velocidades de conducción nerviosa motora y sensitiva, según las técnicas estándares aceptadas de realización, así como en la electromiografía y Ondas F. Los valores de estos resultados se tabularon y capturaron. Con base en las manifestaciones clínicas y características de los estudios electrofisiológicos se compararon con los rangos considerados normales. Se determinó la variante de SGB de cada paciente de acuerdo a los criterios descritos por Asbury y Romper, a la variante axonal motora de SGB se determinó en base a los criterios descritos por Aminoff (Aminoff 1998, Asbury 1990, Ropper 1991). Se capturó la información en un computador con programa de base de datos SPSS 10, efectuando análisis de frecuencias.

3. Resultados preliminares

Hallazgos generales

Revisamos un total de 150 expedientes con diagnóstico de SGB de junio de 1993 a noviembre del 2002. Se excluyeron 18 casos, 13 por no contar con estudios completos y 5 eran formas crónicas de polineuropatía. Incluimos 132 casos, 88 hombres (67%) y 44 mujeres (33%), con una edad promedio de 39 años y con rango de 15 a 80 años. Al comparar el número de pacientes con diagnóstico de SGB (132) con la población de pacientes ingresados a nuestro instituto por cualquier diagnóstico en el periodo de tiempo del estudio (41,615 pacientes), se determinó una prevalencia de 32 casos por 10,000

habitantes. Se encontró reportado un factor desencadenante infeccioso viral probable, en 102 pacientes (77%), infección de vías respiratorias altas en 46%, y un cuadro gastrointestinal en 27%. El tiempo de evolución entre el evento previo infeccioso y el inicio de las manifestaciones fue de 0 a 28 días, con una mediana de 5 días; el 74% de estos pacientes iniciaron con manifestaciones en la primera semana del evento infeccioso. El tiempo de hospitalización promedio fue de 24 días, con un rango de 1 a 185 días. El 76% de los pacientes ameritó al menos tres semanas de internamiento.

Manifestaciones clínicas

Al inicio, el 92% de pacientes (121) manifestaron debilidad, siendo asimétrica en el 14% de ellos. La debilidad fue mayor en miembros inferiores, encontrándose en una gradación menor a 2/5 en 57 pacientes (43%). Los reflejos de estiramiento muscular fueron anormales en todos los pacientes, encontrándose abolidos en el 82% de los casos estudiados. Al ingreso se determinó el diagnóstico en base a criterios clínicos de SGB típico (AIDP- Polineuropatía Desmielinizante Infecciosa Aguda) en 84 pacientes (64%), la variedad Miller-Fisher en 20 de los pacientes (15%), la variante motora pura en 11 (8%), la regional en 5 (4%), mixto en 6 (4.5%) y no se pudieron clasificar por criterios clínicos en 6 pacientes (4.5%). En 127 pacientes (96%) se realizó estudio del LCR, en el 64% de ellos en los primeros 10 días del inicio, se encontró una disociación albúmino-citológica en el 73% de los casos; la celularidad máxima fue de 123 células en un paciente, siendo normales (0 a 5) en el 92%; la proteinorraquia varió de 0 a 965 mg/dl, encontrándose normales en el 42% de los pacientes. Los niveles de glucosa en LCR fueron normales.

Hallazgos neurofisiológicos

En los 132 casos se realizaron estudios de VCN en un promedio de 14 días de iniciado el cuadro clínico de SGB. Se estudiaron los nervios según los criterios de Asbury las alteraciones encontradas fueron disminución de la VCN en 40%, bloqueo de conducción o dispersión temporal en el 89%, latencias dístales prolongadas en el 60%, se encontró alteraciones en la onda "F" con ausencia o prolongación en 81% de los casos, de los cuales el 71% presentaba ausencia de F en por lo menos un nervio, la exploración con electrodo de aguja en la electromiografía mostró denervación activa.

Variantes neurofisiológicas encontradas:

Se estableció el diagnóstico, según los criterios neurofisiológicos de Guillain Barre; 32 casos (24.2%) correspondieron a AIDP, 43 casos (32.6%) a una variante motora [de los cuales 26 casos (19.7%) presentaron la variante AMDN y 17 (12.9%) la variante AMAN], 39 casos (29.5%) correspondieron a variante AMSAN y 18 casos (13.6%) a la variante de Síndrome de Miller Fisher (Figura 1). El diagnóstico clínico fue acorde al reporte neurofisiológico en el 37% de los casos.

4. Discusión

Los hallazgos de las variantes encontradas son muy diferentes a lo ya reportado en otros países, sobre todo los europeos y Estados Unidos. Nuestros resultados arrojan un mayor porcentaje de hombres afectados. Todos los pacientes presentaron alguna alteración en los reflejos, ya sea disminución o abolición, esta última en un 82%. Encontramos bloqueo de conducción o dispersión temporal, seguida de alteraciones en la onda F. Esto guarda relación directa con el daño axonal. La variante más frecuente encontrada fue la AMSAM

con un 30% y si tomamos en cuenta a la variedades motoras de SGB que comprenden AMDN y AMAN tenemos que otro 32 % de nuestra población lo desarrolla con un 62% de compromiso axonal del total de la población estudiada. Corroboramos de esta manera que un importante porcentaje de nuestra población estudiada tiene compromiso axonal, según los estudios de neurofisiológicos, lo cual es mayor de lo reportado en las series de EEUU y Europa, presentando cierta tendencia a un comportamiento semejante a lo reportado en las series de China. La variante de Miller Fisher también fue mayor a lo encontrado en los países occidentales.

5. Conclusiones

Se requiere de la realización de estudios neurofisiológicos para establecer el diagnóstico preciso de las variantes de Guillain Barre. Este estudio demuestra que las variantes del Síndrome de Guillain Barre en México son diferentes a lo reportado por la literatura internacional tradicional de los países occidentales y guardan más similitud a lo reportado en comunidades de China. Las variantes con componente axonal, ya sea motor o sensitivo, son mucho más frecuente de lo que se afirmaba. Al contar con variantes diferentes a las esperadas en nuestro medio, el enfoque diagnóstico deberá ser más rápido y agresivo con una completa evaluación neurofisiológica. El manejo oportuno puede evitar la progresión de la enfermedad, por lo que debe dirigirse el enfoque terapéutico ante la posibilidad de mayor daño axonal en los pacientes, para no incrementar los tiempos de recuperación ante las secuelas y los costos económicos para el estado y la familia en sus cuidados. Estos pacientes por su edad generalmente se encuentran en edad productiva. El SGB no es una enfermedad frecuente, pero, no es una enfermedad rara. Debe ser más estudiada ante el riesgo de mayor morbilidad o secuelas en estos pacientes.

REFERENCIAS

- Alam TA et al. Electrophysiological studies in the Guillain-Barre syndrome: distinguishing subtypes by published criteria. *Muscle Nerve* 1998 Oct;21(10):1275-9.
- Alter M. The epidemiology of Guillain-Barre syndrome. *Annals of Neurology*. 27 Suppl:S7-12, 1990.
- Aminoff Michael J. *Electromyography in clinical practice, clinical and electrodiagnostic aspects of neuromuscular disease*. 3a. edición, Churchill Livingstone, USA 1998. p: 314.
- Asbury A K. *Ann Neurol* 1990; 27 (suppl):S21-S24)
- Asbury AK. Diagnostic considerations in Guillain-Barre syndrome. *Annals of Neurology*. 9 Suppl:1-5, 1981.
- Cheng Q, Jiang G.-X et al. Network members of the Swedish GBS Epidemiology Study Group*. Epidemiological surveillance of Guillain-Barre syndrome in Sweden, 1996-1997. *Acta Neurologica Scandinavica*. 101(2):104-111, February 2000.
- Durand MC et al. Electrophysiologic study of 10 cases of Miller Fisher syndrome. *Rev Neurol (Paris)* 2001 Jan;157(1):72-9.
- Hadden RD et al. Electrophysiological classification of Guillain-Barre syndrome: clinical associations and outcome. Plasma Exchange/Sandoglobulin Guillain-Barre Syndrome Trial Group. *Ann Neurol* 1998 Nov;44(5):780-8.
- Hahn, Angelika F. *Lancet* 1998; 352: 635-41 Guillain- Barre syndrome. *Lancet*. 352(9128):635-641, August 22, 1998.

Kimura Jun. Assessment of individual nerves. En, Electrodiagnosis in diseases of nerve and muscle, principles and practice. Oxford University Pres, New York, 3a. Edición, 2001. Capitulo 6: 130-177.

Kimura J. Proximal versus distal slowing of motor nerve conduction velocity in the Guillain-Barre syndrome. *Annals of Neurology*. 3(4):344-50, 1978 Apr.

Kuwabara S et al. Indicators of rapid clinical recovery in Guillain-Barré syndrome. *J Neurol Neurosurg Psychiatry* 2001;70:560-562 (April).

Marshall E. Bioterror defense initiative injects shot of cash. *Science*. 283(5406):1234-5, 1999 Feb 26.

McKhann G M et al. Acute motor axonal neuropathy: A frequent cause of acute flaccid paralysis in China. *Ann Neurol* 1993; 33; 333-342.

McKhann G M et al. Clinical and electrophysiological aspects of acute paralytic disease of children and young adults in northern China. *The lancet* 1991; 338: 593-97.

Meulstee J. Van der Meché F G A et al. Electrodiagnostic criteria for polyneuropathy and demyelination: application in 135 patients with Guillain-Barré syndrome. *J Neurol Neurosurg Psychiatry* 1995; 59: 482-486

Oomes PG et al. Liver function disturbances in Guillain-Barre syndrome: A prospective longitudinal study in 100 patients. *Neurology*. 46(1):96-100, 1996.

Ropper Allan H. et al. Guillain Barré Syndrome. Contemporary Neurology Series. F.A. Davis Company, Philadelphia, USA 1991.

Van der Meché F G A et al, for the GBS-consensus group of the Dutch Neuromuscular Research Support Centre. Diagnostic and classification criteria for the Guillain-Barré syndrome. *Eur Neurol* 2001; 45: 133-139.

Van Koningsveld R, Van Doorn P A, Schmitz P I M, Ang C W, Van der Meche F G A. Mild forms of Guillain-Barré syndrome in an epidemiologic survey in the Netherlands. *Neurology*. 54(3):620, February 8, 2000.

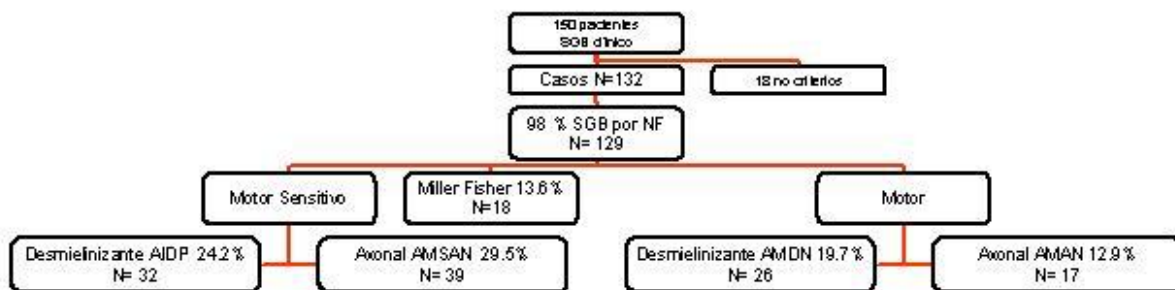


Figura 3. Variantes neurofisiológicas de pacientes con SGB en México.

Factores no Organizacionales y su Posible Relación con Burnout en Personal de Enfermería de Unidad Coronaria

Virginia Cortés Córdova¹, Jorge Elías Torres López^{1,2}, Lucas López Segovia^{3,4}

¹Jefatura de Investigación. Hospital Regional de Alta Especialidad “Dr. Juan Graham Casasús”. Villahermosa, Tab., México.

²División Académica de Ciencias de la Salud. Universidad Juárez Autónoma de Tabasco. Villahermosa, Tab., México. jorge.torres@dacs.ujat.mx

³División Académica de Ciencias Básicas. Universidad Juárez Autónoma de Tabasco. Villahermosa, Tab., México.

⁴Universidad Politécnica de Cataluña. Barcelona, España.

1. Introducción

El Burnout (BO) o *síndrome de estar quemado* se ha descrito principalmente en aquellas personas cuya característica principal es estar en constante contacto con personas (pacientes, clientes, alumnos, etc) (Martínez-López y López-Solache, 2005; Ríos-Risquez et al., 2008; Del Río Moro et al., 2003). La definición más generalizada de BO involucra tres dimensiones: cansancio emocional, despersonalización o cinismo y ausencia o disminución en la realización personal (Maslach et al., 2001; Gil-Monte y Peiró, 1999). Se ha descrito que el BO se asocia con una disminución en el compromiso, desempeño y menor satisfacción laboral (Gutiérrez-Aceves, 2006; Sobrequés et al., 2003, Hernández-Román, 2003) que en su conjunto puede conducir a una mala calidad del servicio que ofrece el profesional a las personas (Palmer-Morales et al., 2005; Sobrequés et al., 2003; Richardsen y Burke, 1991). Los factores de riesgo asociados al desarrollo de BO se pueden agrupar en tres categorías; institucionales, profesionales y características del individuo (Lee y Henderson, 1996). Dentro de éstas últimas se encuentran el estatus demográfico y socioeconómico (Gershon et al., 2007). Tales como el estado civil (soltero, casado, divorciado) e hijos (presencia y número) (Gutiérrez-Aceves, 2006; Palmer-Morales et al., 2005; Martínez-López y López-Solache, 2005). El objetivo del presente trabajo fue determinar el nivel de BO en el equipo de enfermería de la unidad coronaria y su posible relación con el estado civil y presencia de hijos.

2. Materiales y Métodos

Diseño

Se realizó un estudio descriptivo de corte transversal. Se utilizaron dos instrumentos de evaluación autoadministrados. El anonimato y la confidencialidad de los datos se mantuvieron en todo momento.

Población objeto de estudio

La población objeto de estudio estuvo formada por 15 enfermeras(os) que laboran en los 4 diferentes turnos (matutino, vespertino, nocturno y mixto) del servicio de Unidad Coronaria del Hospital de Alta Especialidad “Dr. Juan Graham Casasús”, Villahermosa Tabasco, México. Fueron candidatas a participar en el estudio todos los profesionales de enfermería del servicio de Unidad Coronaria con el único criterio de exclusión de negarse a participar.

Instrumentos de evaluación

Se utilizaron dos instrumentos de evaluación; un cuestionario diseñado *ad hoc* para recoger datos sociodemográficos y laborales, y el otro fue la versión en español del *Maslach Burnout Inventory* (MBI) (Martínez-López y López-Solache, 2005). Este instrumento ha mostrado ser útil, válido y se ha aplicado a personal de enfermería (Martínez-López y López-Solache, 2005; Poncet et al., 2007; Ríos-Risquez, et al., 2008).

Análisis Estadístico

Se realizó estadística descriptiva (promedio y error estándar). También se calculó el puntaje de las tres sub-escalas del MBI y el grado o magnitud de cada una de ellas en base a la siguiente tabla.

| Dimensiones | Bajo | Moderado | Alto |
|----------------------|-----------|----------|-----------|
| Cansancio Emocional | ≤ 16 | 17-26 | > 27 |
| Despersonalización | ≤ 6 | 7-12 | > 13 |
| Realización Personal | ≥ 39 | 32-38 | ≤ 31 |

CE: Cansancio Emocional; DE: Despersonalización; RP: Realización Personal.

Puntos de corte de las dimensiones del MBI descrito por Martínez-López y López-Solache (2005).

Nota: Contrario a CE y DE, alto puntaje en RP indica incremento en la respuesta favorable.

Con el objetivo de analizar e interpretar los resultados de MBI, el grado de BO se clasificó en moderado y alto. Personal con puntaje alto en cansancio emocional o en despersonalización se le consideró con BO moderado, mientras que aquellos que tuvieron puntaje alto en cansancio emocional, puntaje alto en despersonalización y puntaje bajo en realización personal se les consideró BO alto (Goebring et al., 2005). También se calculó en porcentaje de enfermeras que tuvieron alterado al menos una de las dimensiones del MBI, las que presentaron 2 y las que presentaron 3 (Molina-Siguero et al., 2003).

3. Resultados

Se encuestó al 100% del personal de enfermería del servicio de unidad coronaria. La edad promedio fue 36.27 ± 2.11 años, la mínima fue de 24 y la máxima 53 años. El 86.7 % corresponde al género femenino con edad de 37.9 ± 2.1 y la edad de los hombres 25.5 ± 1.5 . La frecuencia con la que se presentó el cansancio emocional y despersonalización fue 33.3% y 20% respectivamente. Mientras que más de la mitad presentó baja realización personal (Tabla 1). La alteración de las tres dimensiones del BO es más frecuente en las casadas (Tabla 1).

Tabla 1. Magnitud de las dimensiones del Síndrome de Burnout respecto a la variable no organizacional estado civil.

| Estado Civil | Cansancio Emocional (%) | | Despersonalización (%) | | | Realización Personal (%) | | |
|--------------|-------------------------|-------------|------------------------|-------------|-------------|--------------------------|-------------|-------------|
| | Bajo | Moderado | Bajo | Moderado | Alto | Alta | Moderada | Bajo |
| Casada | 40.0 | 20.0 | 46.7 | 13.3 | 0 | 20.0 | 13.3 | 26.7 |
| Soltera | 6.67 | 0 | 0 | 0 | 6.67 | 0 | 6.67 | 0 |
| Unión Libre | 13.3 | 6.67 | 20.0 | 0 | 0 | 6.67 | 0 | 13.3 |
| Divorciada | 0 | 6.67 | 6.67 | 0 | 0 | 0 | 0 | 6.7 |
| Viuda | 6.67 | 0 | 6.67 | 0 | 0 | 0 | 0 | 6.7 |
| Total | 66.6 | 33.3 | 80.0 | 13.3 | 6.67 | 26.7 | 20.0 | 53.3 |

El 20% (2 hombres y 1 mujer) no mostraron alguna alteración en las dimensiones del BO (grupo sano) (Tabla 2). El 80% de la muestra presentó al menos una dimensión alterada (moderado o alto) (Tabla 2). El promedio de edad obtenido por el grupo que no presento alteraciones fue de 29.33 ± 3.93 , seguido del grupo con una alteración el cual presento un promedio de 34.20 ± 2.27 años, el grupo que presento dos dimensiones alteradas obtuvo un promedio de 40.71 ± 3.21 (Tabla 2). De acuerdo al estado civil, las solteras tienen una mayor frecuencia en el grupo sano obteniendo cerca del 70% en este, sin embargo representan más de la mitad en las dimensiones con una y dos alteraciones; la variable unión libre obtuvo una frecuencia de 33% sin alteración en ninguna de las dimensiones de Burnout y 34% distribuido en los grupos con una y dos dimensiones alteradas, la presencia de soltera y divorciada ocurrió solo en el grupo de dos dimensiones alteradas con una frecuencia de 14% en ambas, del mismo modo se presentó la variable viuda en el grupo de una dimensión con frecuencia de 20% (Tabla 2).

Tabla 2. Factores asociados a las características individuales en proporción al número de dimensiones alteradas.

| | No. Dimensiones de BO Alteradas | | |
|--|---------------------------------|------------------|------------------|
| | 0 (20%) n=3 | 1 (33%) n=5 | 2 (47%) n=7 |
| Edad Promedio \pm E.E. | 29.33 \pm 3.93 | 34.20 \pm 2.27 | 40.71 \pm 3.21 |
| Estado Civil | | | |
| Casada | 67% | 60% | 57% |
| Unión Libre | 33% | 20% | 14% |
| Viuda | | 20% | |
| Divorciada | | | 14% |
| Soltera | | | 14% |
| Hijos | | | |
| 1 | 33.3% | 40% | 14% |
| 2 | 33.3% | 20% | 14% |
| 3 | | 20% | 43% |
| 4 | | | 14% |
| 0 | 33.3% | 20% | 14% |

Respecto a la población que presentó al menos una dimensión alterada, más del 40% presentó cansancio emocional moderado, cerca del 20% y más del 8% nivel moderado y alto en la dimensión de despersonalización, respectivamente (Tabla 3). 20% alcanzó nivel moderado y cerca del 70% niveles bajos en la realización personal. (Tabla 3)

Tabla 3. Niveles de Burnout: Se evaluaron en las tres dimensiones con su correspondiente puntaje en el personal que al menos tiene una dimensión alterada.

| DIMENSIONES | Bajo (%) | Moderado (%) | Alto (%) | Puntaje Promedio \pm e.e |
|----------------------|----------|--------------|----------|----------------------------|
| CANSANCIO EMOCIONAL | 58.3 | 41.7 | --- | 16 \pm 1.0 |
| DESPERSONALIZACIÓN | 75 | 17 | 8.3 | 4.9 \pm 1.4 |
| REALIZACIÓN PERSONAL | 67 | 25 | 8.3 | 30 \pm 1.6 |

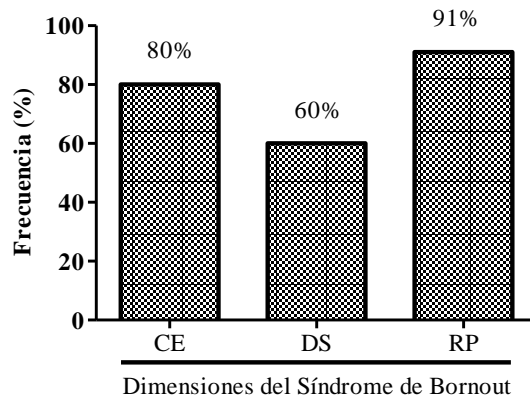


Figura 1. Porcentaje del personal que tiene al menos 1 hijo y que tiene al menos una dimensión del BO afectada..

4. Discusión

En el presente estudio se observó que la edad promedio del personal de enfermería del servicio de unidad coronaria fue alrededor de 36 años. Lo anterior concuerda con lo descrito en la literatura con respecto a otros servicios de cardiología (Molina-Siguero et al., 2003; Del Río Moro et al., 2003) y con servicios de cuidados intensivos (Poncet et al., 2006). En nuestro estudio se observó baja frecuencia de CE y DS así como un alto porcentaje en la presencia de realización personal disminuida lo cual difiere de otros trabajos (Ríos-Risquez et al., 2007; Poncet et al., 2006; Grau et al., 2005)

Se han descrito diversos factores de riesgo para el desarrollo de Burnout, entre ellas, el estado civil, en el presente estudio se identificó que las enfermeras casadas presentan más alteraciones en las dimensiones del BO, principalmente en la realización personal disminuida. Lo anterior coincide con lo descrito por otros autores (Palmer et al., 2005; Martínez-López y López Solache, 2005). Una posible explicación es el rol femenino en el hogar (Gil-Monte, 2002), sin embargo, otros autores sugieren que la soltería puede ser un desencadenante del síndrome de BO (Gutiérrez-Aceves et al., 2006; Hidalgo et al., 1994; Guevara, Henao, Herrera, 2004).

Otro de los factores es los hijos, la literatura sugiere que la presencia y número de hijos representa un factor de riesgo para presentar Burnout (Gutiérrez-Aceves et al., 2006). Sin embargo, en nuestro estudio la presencia o no y el número de hijos no parece ser un factor que esté involucrado con el con el número alterado de dimensiones del BO.

La edad se ha correlacionado a la presencia de Burnout, asociándolo en las personas que se encuentran entre 30 y 40 años (Sobreque et al., 2003; Palmer et al., 2005; Martínez-López y López Solache, 2005), otras literaturas definen la edad joven como factor negativo debido al proceso de adaptación al ejercicio profesional (Gutiérrez-Aceves et al., 2006; Moreno et al. 1996), en nuestra población se presentaron diferencias significativas en el promedio de edad entre el grupo de dos dimensiones alteradas y el grupo sano, con promedio de 40.7 ± 3.21 años y promedio de 29.33 ± 3.93 años respectivamente, lo que podría indicar que a mayor edad, mayor probabilidad de presentar Burnout.

REFERENCIAS

- Del Río Moro, O., Perezagua García MC., Vidal Gómez, B. (2003). Síndrome de Burnout en los enfermeros/as del hospital virgen de la salud de Toledo. *Enfermería en Cardiología* 28, 24-29.
- Gershon, R. M. R., Stone, W. P., Zeltser, M., Faucett, J., Macdavitt, K., Chou, S.S. (2007). Organizational Climate and Nurse Health Outcomes in the United States: A Systematic Review. *Industrial Health* 45, 622-636.
- Gil-Monte P R. y Peiró J M. (1999). Validez Factorial de Maslach Burnout Inventory en una Muestra Multiocupacional. *Psicothema* 11, 679-689.
- Goebring, C., Bouvier Gallacchi, M., Kiinzi, B., Bovier, P. (2005). Psychosocial and Professional Characteristics of Burnout in Swiss Primary Care Practitioners: a Cross-Sectional Survey. *Swiss Med Wkly* 135, 101-108.
- Grau, A., Suñer, R., García, M M. (2005). Desgaste profesional en el personal sanitario y su relación con los factores personales y ambientales. *Gaceta Sanitaria* 19, 463-70.
- Guevara, CA., Henao, DP., Herrera, JA.(2004). Síndrome de desgaste profesional en médicos internos y residentes. *Colombia Médica* 35, 173-8.
- Gutiérrez Aceves, G A., Celis López, M A., Moreno Jiménez S., Farias Serratos, F., Suárez Campos, J J. (2006). Síndrome de Burnout. *Arch Neurocién*, 11, 305-309.
- Hernández J R. (2003). Estrés y burnout en profesionales de la salud de los niveles primario y secundario de atención. *Revista Cubana Salud Publica*. 29, 103-110.
- Hidalgo, RI., Díaz, GRJ. (1994). Estudio de la Influencia de los factores laborales y sociodemográficos en el desarrollo del síndrome de agotamiento profesional en el área de medicina de especialidades de INSALUD de Ávila. *Medicina clínica* 103, 408-412.
- Lee, V., Henderson, M C.(1996). Occupational Stress and organizational commitment in nurse administrators. *Journal Nurs Administ* 26, 21-8
- Martínez-López, C., López-Solache, G. (2005). Características del Síndrome de Burnout en un Grupo de Enfermeras Mexicanas. *Archivos en Medicina Familiar* 7, 6-9.
- Maslach, C., Schaufeli, W B., Leiter, M P. (2001). Job burnout. *Annu Review Psychol* 52, 397-422.
- Molina Siguero, A., García Pérez, M A., González, M A., y Cermeño, P C. (2003). Prevalence of Worker Burnout and psychiatric Illness in Primary Care Physicians in a Health Care Area in Madrid. *Atención Primaria* 31, 564-571.
- Moreno, J B., Peñalcoba, P C. (1996). *Manual de Psicología de la Salud*. Madrid: Siglo XXI.
- Palmer Morales, LY., Gómez Vera, A., Cabrera Pivaral, C., Price Velaz, R., Searcy Bernal, R. (2005). Prevalencia del síndrome de agotamiento profesional en médicos anesestesiólogos de la ciudad de Mexicali. *Gaceta Médica de México* 141,181-3.
- Palmer, Y., Gómez-Vera, A., Cabrera Pivaral, C., Prince-Vélez, R., Searcy, R. (2005). Factores de Riesgo Organizacionales Asociados al Síndrome de Burnout en Médicos Anesestesiólogos. *Salud Mental* 28, 83-9.
- Poncet, C M., Toullic, P., Papazian, L., Ketish-Barnes, Timsit, J F., Pochard, F., Chevret, S., Schlemmer, B. and Azoulay, É. (2007). Burnout Syndrome in Critical Care Nursing Staff. *American Journal Of Respiratory and Critical Care Medicine* 175, 698-704.
- Richardsen, A M., Burke, R J. (1991). Occupational stress and job satisfaction among physicians: sex differences. *Society SCI medical* 33, 1179-87.
- Ríos Riskey, M I., Godoy Fernández, C., Peñalver Hernández, F., Alonso Tovar, A.R., López Alcaraz, F., López Romera, A., Garnéz González, S., Salmerón Saura, E., Lopez

Real, M D., Ruiz Sánchez, R., Simón Domingo, P., Manzanera Nicolás, J L., Menchón Almagro, M A. y Liébanas Bellón. (2008). Estudio Comparativo del Burnout en el Personal de Enfermería de Cuidados Intensivos y Urgencias. *Enfermería Intensiva* 19, 2-13.

Sobrequés, J., Cebria, J., Segura, J., García, M., Juncosa, S. (2003). La satisfacción laboral y el desgaste profesional de los médicos de atención primaria. *Atención Primaria* 31, 227-233.

Estudio Inmunohistoquímico de los Variables Categóricas Ordinales por Caspasa 3 y Alfa-Fetoproteína de las Lesiones Hepáticas Producidas por el Virus de la Hepatitis C y su Relación con el Hepatocarcinoma. Captación Genómica por Fusión Celular: Una Teoría de la Involución de las Especies

Leonardo Varela Rodríguez¹, Martín Varela Vindas²

¹Universidad Autónoma de Centroamérica

²Universidad de Salamanca

Resumen

Medición de la actividad de los marcadores inmunohistoquímicos caspasa 3 y alfa-fetoproteína en biopsias hepáticas con los diagnósticos de hepatitis crónica y cirrosis hepática por el virus de la hepatitis C y el hepatocarcinoma en general.

1. Introducción

El hígado es un órgano de origen endodérmico muy complejo. El carcinoma (1), (2), (5), (7), (8), (9), (10) hepatocelular es un tumor primario del hígado y es el quinto tumor maligno más frecuente del mundo, y es la tercera causa más frecuente de muerte por cáncer a nivel mundial. (11), (12), (13), (18), (20). Por tanto este estudio busca nuevos indicios etiopatogénicos de la hepatocarcinogénesis, que brinden nuevas estrategias diagnósticas y terapéuticas para un cáncer que provoca seiscientos mil muertes al año alrededor del mundo. (21), (29), (30), (31), (32).

2. Metodología

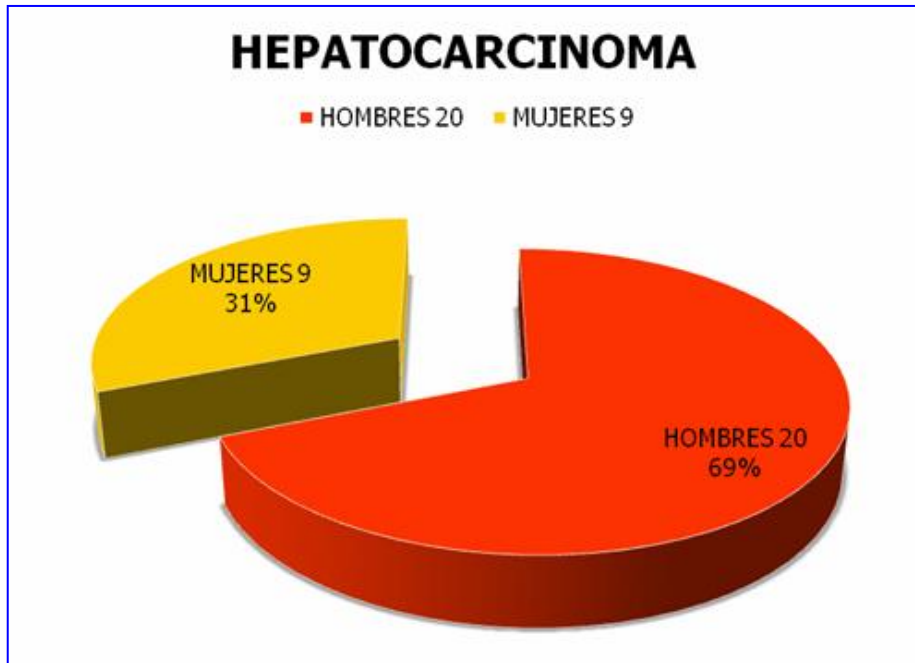
- Tipo de estudio: es un estudio transversal analítico descriptivo para determinar una variable como factor diagnóstico, pronóstico y de riesgo, para el desarrollo de hepatocarcinoma que es el estándar áureo de referencia. Se analiza la prevalencia y distribución de variables categóricas ordinales, por medio de un sistema inmunohistoquímico, a través del desenmascaramiento antigénico de la peroxidasa.
- Muestreo y tamaño de la muestra: 1768 muestras para biopsia hepática del archivo computarizado de Anatomía Patológica, recolectados del 1 de diciembre de 1997 al 30 de noviembre de 2007. Todas las biopsias por cirrosis hepática a Virus C (14 casos), 1 de cada cuatro biopsias de 125 casos de Hepatitis crónica a Virus C (31 casos) por ordenador. Además todos los casos de hepatocarcinoma (29 casos) por diversa etiología, del periodo 1997-2007, del Hospital Universitario de Salamanca, España.

3. Resultados

La distribución por sexo en el Hepatocarcinoma (HCA) es: 20 hombres (69%) y 9 mujeres (31%).

La caspasa 3 positiva en 3%, 7% y 69% de los grupos de hepatitis crónica Virus C, cirrosis virus C y hepatocarcinoma, respectivamente. P significativa (<0.001) sólo en el grupo caspasa 3.

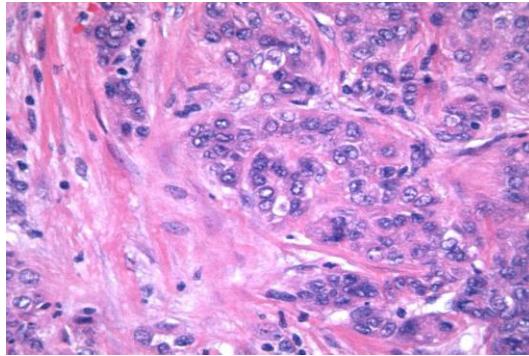
FIG. 1- DIAGRAMA DE SECTORES POR FRECUENCIA DE GÉNERO



Fuente: Bioestadística del Servicio de Anatomía Patológica. Hospital Universitario de Salamanca. Dic. 1997-
Nov 2007.

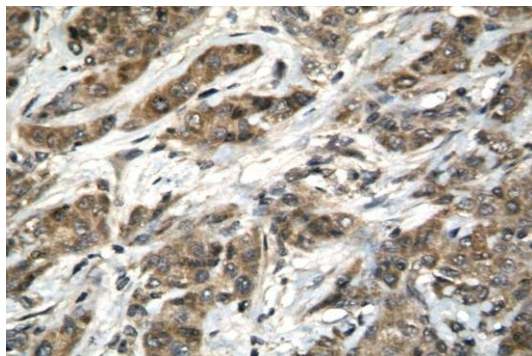
FOTOS DE BIOPSIAS HEPÁTICAS EN HEPATOCARCINOMA (HCA)

1. Hematoxilina eosina HCA:



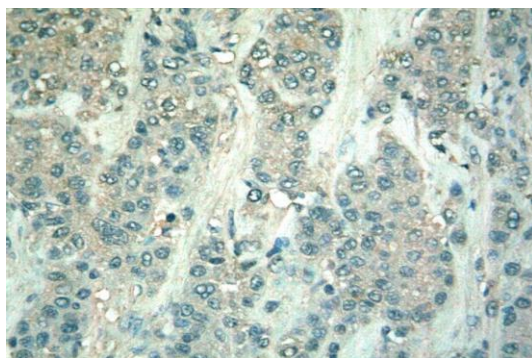
Hematoxilina eosina
Hepatocarcinoma 40x
HCA: Hepatocarcinoma

2. α -fetoproteína HCA.



α -fetoproteína 40x
HCA: Hepatocarcinoma

3. Caspasa3 HCA.



Caspasa3 40x
HCA: Hepatocarcinoma

TABLAS DE CONTINGENCIA

RESULTADOS DEL ANÁLISIS ESTADÍSTICO.

Frecuencia y distribución absoluta y porcentual de PIH

Tabla.1

Actividad de α -fetoproteína y caspasa3, en biopsias con diagnósticos de Hepatocarcinoma y cirrosis VHC.

| 43 Biopsias | 29 HCA | | | | ATC | | | | | 14 Cirrosis-VHC | | |
|-----------------------|--------|----|----|-------------|---------|-----|------|------|-----|-----------------|----|-------------|
| | + | ++ | - | % positivos | P | S | E | VPP | VPN | + | - | % positivos |
| α fetoproteína | 16 | 2 | 11 | 62% | ns | 62% | 100% | 100% | 56% | 0 | 14 | 0 |
| Caspasa 3 | 18 | 2 | 9 | 69% | < 0.001 | 69% | 92% | 92% | 59% | 1 | 13 | 7% |

Frecuencia y distribución absoluta y porcentual de PIH *Tabla.1'*

Actividad de α fetoproteína y caspasa3, en biopsias con diagnósticos de Hepatocarcinoma y Hepatitis VHC.

| 60 Biopsias | 29 HCA | | | | ATC | | | | | 31 Hepatitis-VHC | | |
|-----------------------|--------|----|----|-------------|---------|-----|-----|-----|-----|------------------|----|-------------|
| PIH | + | ++ | - | % positivos | P | S | E | VPP | VPN | + | - | % positivos |
| α fetoproteína | 16 | 2 | 11 | 62% | ns | 62% | 96% | 94% | 73% | 1 | 30 | 3% |
| Caspasa 3 | 18 | 2 | 9 | 69% | < 0.001 | 69% | 96% | 95% | 76% | 1 | 30 | 3% |

PIH = Pruebas Inmunohistoquímicas; Actividad: Leve = +; Moderada = ++; Inactiva = -; HCA = Hepatocarcinoma; VHC = virus hepatitis C; ATC = Análisis de Tablas de Contingencia; P = Probabilidad (χ^2 Pearson); I = Incalculable; S = Sensibilidad; E = Especificidad; VPP = Valor Predictivo Positivo; VPN = Valor Predictivo Negativo.

Fuente. Muestras de biopsia del Servicio de Anatomía Patológica del Hospital Universitario de Salamanca. Dic. 1997 - Nov. 2007

Frecuencia y distribución relacionada de resultados de pruebas inmunohistoquímicas en HCA *Tabla 2*

| 29 Hepatocarcinomas | 4 HCA | 12 HCA | 8 HCA | 5 HCA | Positividad |
|---------------------------|-------|--------|-------|-------|-------------|
| <i>c 3</i> | - | + | + | - | 20 / 69 % |
| <i>Alfafetoproteína +</i> | 3 | 5 | 6 | 4 | 18 / 62 % |
| <i>Alfafetoproteína -</i> | 1 | 7 | 2 | 1 | 11 / 38 % |
| Total | 4 | 12 | 8 | 5 | 29 / 100 % |

Positivo = + ; Negativo = -; HCA = Hepatocarcinoma; Caspasa 3 = c 3

Fuente. Muestras de biopsia del Servicio de Anatomía Patológica del Hospital Universitario de Salamanca. Dic. 1997 - Nov. 2007.

4. Discusión

El hepatocito se infecta con el virus de la hepatitis C y posteriormente se produce una nueva célula hepatocitaria al fusionarse la célula hepática senescente (24) con la célula madre hepática o con la célula precursora hepática (oval). (4), (23). La denominaremos: célula simbiooncomutagénica hepática, la cual luego de reproducirse aceleradamente por el influjo viral (22) y en un contexto anatomopatológico de fibrosis inflamatoria, que puede corresponder a hepatitis crónica o a cirrosis, la misma presenta escape apoptósico y al salvar la apoptosis se sigue reproduciendo gracias a que las células inmunes también presentan agotamiento celular. Al activarse las vías oncogénicas y del desarrollo tumoral respectivamente se produce una catástrofe simbiooncomutagénica. (16), (26), (28). Esta nueva célula bajo el constante influjo de marcadores de superficie y producción exagerada de telomerasa desembocan en la producción o génesis del hepatocarcinoma y células madre

acompañantes que derivan en células con capacidad de desarrollo en tejidos aberrantes incluyendo vasos de neoformación para el crecimiento tumoral y más adelante el desarrollo de metástasis. Telomerasa y telómeros tumorales alcanzan la inmortalidad del hepatocarcinoma. (3), (33).

La telomerasa es la máquina del tiempo pero con dos túneles del campo unificado en un solo plano: la inmortalidad. El primer túnel es la vida como la conocemos. El segundo túnel es la neoplasia (muerte humana). Las células fetales, embrionarias y germinales, así como las neoplásicas expresan telomerasa. Recientes investigaciones muestran la importancia relevante entre telomerasa y cáncer, células madre, envejecimiento y síndromes hereditarios. (14). ¿Es una relación causa-efecto? Nuestra hipótesis es que sí. Primero la telomerasa y luego el tumor en nuestro modelo biomédico de hepatitis crónica y cirrosis hepática asociada a virus C (25) con posterior evolución a hepatocarcinoma. (15), (27). La caspasa 3 presente en el 69% de los casos de hepatocarcinoma nos demuestra la captación genómica por fusión celular ($p < 0.001$) asociada a vías de señalización (incluyendo mutación de vías de supresión tumoral y activación de vías oncogénicas) y la presencia del VHC, con posterior escape caótico apoptótico (17) de la célula simbiocomutagénica hepática por fusión celular derivados a través del segundo túnel neoplásico al fracasar la inmunovigilancia de un sistema ordenado (19), es lo que llamaremos el EFECTO GALAPAGOS (túnel caótico de un sistema ordenado) en contra posición del efecto mariposa (sistema caótico y aparición de ventana ordenada).

5. Conclusiones

1. Predominio de pacientes del sexo masculino en los tres grupos.
2. La caspasa 3 se encuentra activa en un 69% de los casos del grupo Hepatocarcinoma, versus 7% y 3% de los grupos cirrosis y hepatitis crónica a VHC respectivamente. Esta gran diferencia ($p < 0.001$) nos demuestra que el parénquima hepático tumoral del Hepatocarcinoma está produciendo la caspasa 3 por captación genómica heredada por fusión celular en el proceso de simbiocomutagénesis de la teoría de la fusión celular, de células hepáticas senescentes con células madre hepáticas y ovales, bajo múltiples vías de señalización y estrés oxidativo por inducción del virus de la hepatitis C, esto lo denominamos el EFECTO GALÁPAGOS.
3. Se recomienda ampliar el estudio con antígenos de superficie de la célula hepatocitaria tumoral para aclarar su origen jerárquico Bi ó multipartito por reacción en cadena de la polimerasa (P.C.R), según la teoría de la fusión celular en la simbiocomutagénesis del Hepatocarcinoma y VHC.
4. Se plantea el cuarto postulado de la TEORÍA CELULAR: "En circunstancias especiales una nueva célula se origina por fusión celular de dos o más células preexistentes".

REFERENCIAS

- Allende Daniela S and Yerian Lisa. Immunohistochemical Markers in the Diagnosis of Hepatocellular Carcinoma. Pathology Case Reviews 2009; 14: 40-46.
- ARMSTRONG GL, ALTER MJ, MCQUILLAN GM, MARGOLIS HS. The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. Hepatology, 2000; 31: 777-782.
- BLACKBURN, EH. Switching and signaling at the telomere. Cell, 2001; 106: 661-673.

Bosch Barrera J, Lopez – Picazo Gonzalez J.M., Garcia – Foncillas Lopéz J, Prósper Cardoso F. Células Madre y Cáncer: dilucidando el origen de la célula madre tumoral. *Rev Med Univ Navarra/Vol 51, N°2, 2007, 14 – 17.*

CAPOCACCIA RICCARDO, Sant Milena et al. Hepatocellular carcinoma: trends of incidence and survival in Europe and the United States at the end of the 20th century. *AM.J. Gastroenterol, 2007; 102: 1661-1670.*

CHOO Q-L, KUO G, WEINER AM, JONES DB, et al. Isolation of Ac DNA clone derived from A blood-borne non-A, non-B viral hepatitis genome. *Science, 1989; 244: 359-362.*

DAVILA JA, MORGAN RO, SHAIB Y, MCGLYNN KA, EL-SERAG HB. Hepatitis C infection and the increasing incidence of hepatocellular carcinoma: a population-based study. *Gastroenterology, 2004; 127: 1372-1380.*

EL-SERAG HB, DAVILA JA, PETERSEN NJ, MCGLYNN KA. The continuing increase in the incidence of hepatocellular carcinoma in the United States: an update. *Ann Intern Med, 2003; 139: 817-823.*

EL-SERAG HB, MASON AC. Rising incidence of hepatocellular carcinoma in the United States. *N Engl J Med, 1999; 340:745-750.*

EL-SERAG HB, MASON AC. Risk factors for the rising rates of primary liver cancer in the United States. *Arch Intern Med, 2000; 160: 3227-3230.*

EL-SERAG HB, RUDOLPH KL. Hepatocellular carcinoma: Epidemiology and Molecular Carcinogenesis. *Gastroenterology, 2007; 132: 2557-2576.*

EL-SERAG HB. Hepatocellular carcinoma: recent trends in the United States. *Gastroenterology, 2004; 127 (5 suppl 1): S27-S34.*

FAUCI, Anthony S; BRAUNWALD, Eugene, et al. *Harrison. Principios de Medicina Interna. McGraw Hill, 17ª Edición. 2008.*

FIBBE WILLEM, E. Telomerase mutatis in aplastic anemia. Editorials. *N Engl J Med, 2005; 352: 1481-1482*

FRÖHLING, Stefan; DÖHNER, Hartmut. Chromosomal abnormalities in cancer. *New England. J Med, 2008 August 14; 359: 722-734.*

HAYFLICK, L. The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res, 1965; 37: 614-636.*

LANED P. p53: guardian of the genome. *Nature, 1992; 358: 1516.*

LLOVET, et al. Presentación oral, 2007. Annual Meeting, 4 junio 2007. Chicago, IL, USA.

MITCHELL, Peter. Acoplamiento de la fosforilación a la transferencia de electrones e hidrógeno mediante un mecanismo de tipo quimiosmótico. *Nature, 1961; Volume 191: 144-148.*

NEAL KEITH, R. On Behalf of the Trent Hepatitis C Study Group. *GUT 2007; 56: 1098-1104.*

PARKIN DM, BRAY F, et al. Estimating the world cancer burden: Globocan 2000. *Int J Cancer 2001; 94:153-156.*

PAZ BOUZA, José Ignacio. Aspectos histopatológicos de diferentes órganos del ratón inoculados con virus herpes simples de tipo I y II. Tesis Doctoral. Facultad de Medicina. Universidad de Salamanca. España. 1978.

POTENZA, Leonardo; LUPPI, Mario; BAROZZI, Patrizia; et al. HHV- GA syncytial giant-cell hepatitis. *New England. J Med, Number 6. 2008 August 7; 359: 593-602.*

ROSENTHAL, Nadia. Prometheus's Vulture and stem-cell promise. Number 3. *New England. Journal of Medicine, 2003 July 17; 349: 267-274.*

- RUDOLPH KL, CHANG S, MILLARD M, SCHREIBER-AGUS N, DEPINHO RA. Inhibition of experimental liver cirrhosis in mice by telomerase gene delivery. *Science*, 2000; 287: 1253-1258.
- SÁNCHEZ POSADA, Raúl. Características biológicas de la célula en el cáncer de colon con metástasis ganglionares. Tesis Doctoral. Salamanca, 2007.
- SUN WEI, XING BAOCAI et al. Proteome analysis of hepatocellular carcinoma by two-dimensional difference gel electrophoresis. *Molecular and Cellular proteomics*, 2007; 6: 1798-1808.
- TANG, Weiliang; LAZARO, Catherine A; et al. Responses of non transformed human hepatocytes to conditional expression of full-length hepatitis C virus open reading frame. *Am J Pathol*, 2007; 171: 1831-1846.
- WONG JB, MCQUILLAN GM, MCHUTCHISON JG, POYNARD T. Estimating future hepatitis C morbidity, mortality, and costs in the United States. *Am J Public Health*, 2000; 90: 1562-1569.
- YOSHIZAWA H. Hepatocellular carcinoma associated with hepatitis C virus infection in Japan: projection to other countries in the foreseeable future. *Oncology*, 2002; 62 (suppl 1): 8-17.
- YU MW, CHEN CJ. Elevated serum testosterone levels and risk of hepatocellular carcinoma. *Cancer Res*, 1993; 53: 790-794.
- YU MW, YANG YC, YANG SY, et al. Hormonal markers and hepatitis B virus-related hepatocellular carcinoma risk: a nested case-control study among men. *J Natl Cancer Inst*, 2001; 93: 1644-1651.
- ZEN YOH, FUJII TAKAHIKO et al. Histological and culture studies respect to ABC62 expression support the existence of a cancer cell hierarchy in human hepatocellular carcinoma. *AJP*, 2007; 170: 1750-1762.

Estudio Inmunoquímico de los Variables Categóricas Ordinales por P53 y Alfa-Fetoproteína de las Lesiones Hepáticas Producidas por el Virus de la Hepatitis C (VHC) y su Relación con el Hepatocarcinoma (HCA)

Leonardo Varela Rodríguez¹, Martín Varela Vindas²

¹Universidad Autónoma de Centroamérica

²Universidad de Salamanca

Resumen

Medición de la actividad de los marcadores inmunoquímicos de p53 y alfa-fetoproteína en biopsias hepáticas con los diagnósticos de hepatitis crónica y cirrosis hepática por el virus de la hepatitis C y el hepatocarcinoma en general.

1. Introducción

El hígado es un órgano de origen endodérmico muy complejo. El carcinoma (1), (2), (5), (7), (8), (9), (10) hepatocelular es un tumor primario del hígado y es el quinto tumor maligno más frecuente del mundo, y es la tercera causa más frecuente de muerte por cáncer a nivel mundial. (11), (12), (13), (18), (20). Por tanto este estudio busca nuevos indicios etiopatogénicos de la hepatocarcinogénesis, que brinden nuevas estrategias diagnósticas y terapéuticas para un cáncer que provoca seiscientos mil muertes al año alrededor del mundo. (21), (29), (30), (31), (32).

2. Metodología

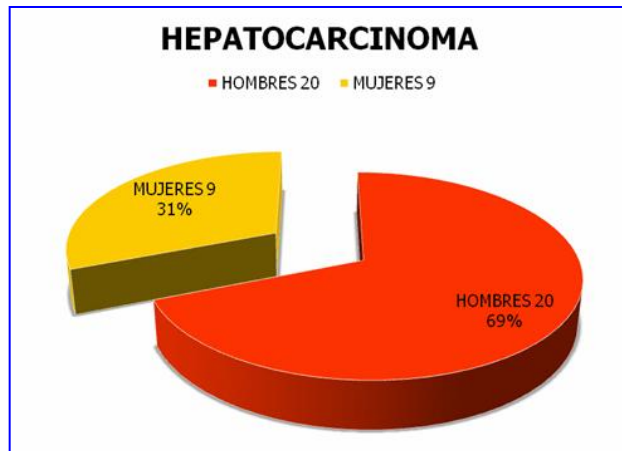
- Tipo de estudio: es un estudio transversal analítico descriptivo para determinar una variable como factor diagnóstico, pronóstico y de riesgo, para el desarrollo de hepatocarcinoma que es el estándar áureo de referencia. Se analizará la prevalencia y distribución de variables categóricas ordinales, por medio de un sistema inmunoquímico, a través del desenmascaramiento antigénico de la peroxidasa.
- Muestreo y tamaño de la muestra: 1768 muestras para biopsia hepática del archivo computarizado de Anatomía Patológica, recolectados del 1 de diciembre de 1997 al 30 de noviembre de 2007. Todas las biopsias por cirrosis hepática a Virus C (14 casos), 1 de cada cuatro biopsias de 125 casos de Hepatitis crónica a Virus C (31 casos) por ordenador. Además todos los casos de hepatocarcinoma (29 casos) por diversa etiología, del periodo 1997-2007, del Hospital Universitario de Salamanca.

3. Resultados

La distribución por sexo en el Hepatocarcinoma (HCA) es: 20 hombres (69%) y 9 mujeres (31%).

La p53 es positiva en 3%, 29%, 41% y alfafetoproteína positiva en 3%, 0%, 62%, de los grupos de hepatitis crónica Virus C, cirrosis virus C y hepatocarcinoma, respectivamente.

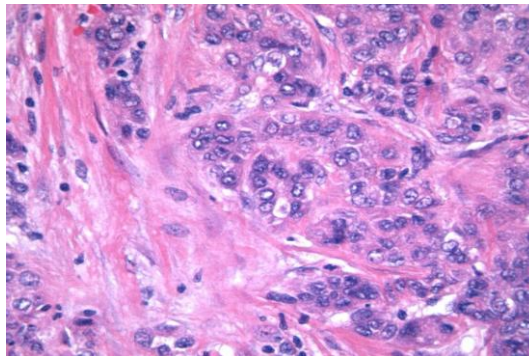
FIG. 1- DIAGRAMA DE SECTORES POR FRECUENCIA DE GÉNERO



Fuente: Bioestadística del Servicio de Anatomía Patológica. Hospital Universitario de Salamanca. Dic. 1997-
Nov 2007.

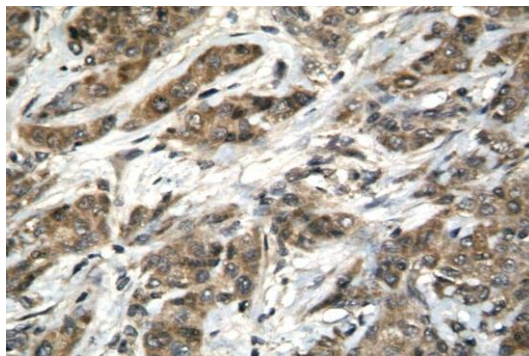
FOTOS DE BIOPSIAS HEPÁTICAS EN HEPATOCARCINOMA (HCA)

4. Hematoxilina eosina HCA:



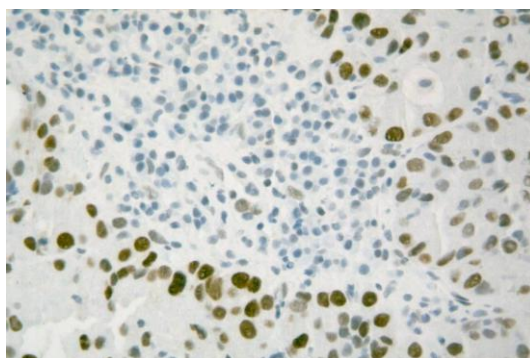
Hematoxilina eosina
Hepatocarcinoma 40x
HCA: Hepatocarcinoma

5. α -fetoproteína HCA.



α -fetoproteína 40x
HCA: Hepatocarcinoma

6. p53 HCA.



p53 40x

HCA: Hepatocarcinoma

**TABLAS DE CONTINGENCIA
RESULTADOS DEL ANÁLISIS ESTADÍSTICO.**

Frecuencia y distribución absoluta y porcentual de PIH *Tabla.1*

Actividad de p53, α fetoproteína en biopsias con diagnósticos de Hepatocarcinoma y cirrosis VHC.

| 43 Biopsias | 29 HCA | | | | ATC | | | | | 14 Cirrosis-VHC | | |
|-----------------------|--------|----|----|-------------|-----|-----|------|------|-----|-----------------|----|-------------|
| | + | ++ | - | % positivos | P | S | E | VPP | VPN | + | - | % positivos |
| p53 | 11 | 1 | 17 | 41% | ns | 41% | 71% | 75% | 37% | 4 | 10 | 29% |
| α fetoproteína | 16 | 2 | 11 | 62% | ns | 62% | 100% | 100% | 56% | 0 | 14 | 0 |

Frecuencia y distribución absoluta y porcentual de PIH *Tabla.1'*

Actividad de p53, α fetoproteína en biopsias con diagnósticos de Hepatocarcinoma y Hepatitis VHC.

| 60 Biopsias | 29 HCA | | | | ATC | | | | | 31 Hepatitis-VHC | | |
|-----------------------|--------|----|----|-------------|-----|-----|-----|-----|-----|------------------|----|-------------|
| | + | ++ | - | % positivos | P | S | E | VPP | VPN | + | - | % positivos |
| p53 | 11 | 1 | 17 | 41% | ns | 41% | 96% | 92% | 63% | 1 | 30 | 3% |
| α fetoproteína | 16 | 2 | 11 | 62% | ns | 62% | 96% | 94% | 73% | 1 | 30 | 3% |

PIH = Pruebas Inmunohistoquímicas; Actividad: Leve = +; Moderada = ++; Inactiva = -; HCA = Hepatocarcinoma; VHC = virus hepatitis C; ATC = Análisis de Tablas de Contingencia; P = Probabilidad

(χ^2 Pearson); I = Incalculable; S = Sensibilidad; E = Especificidad; VPP = Valor Predictivo Positivo; VPN = Valor Predictivo Negativo.

Fuente. Muestras de biopsia del Servicio de Anatomía Patológica del Hospital Universitario de Salamanca. Dic. 1997 - Nov. 2007

Frecuencia y distribución relacionada de resultados de pruebas *Tabla2.* inmunohistoquímicas en HCA

| 29 Hepatocarcinomas | 4 HCA | 12 HCA | 8 HCA | 5 HCA | Positividad |
|---------------------|-------|--------|-------|-------|-------------|
| p53 | + | - | + | - | 12 / 41 % |
| Alfafetoproteína + | 3 | 5 | 6 | 4 | 18 / 62 % |
| Alfafetoproteína - | 1 | 7 | 2 | 1 | 11 / 38 % |
| Total | 4 | 12 | 8 | 5 | 29 / 100 % |

Positivo = + ; Negativo = -; HCA = Hepatocarcinoma.

Fuente. Muestras de biopsia del Servicio de Anatomía Patológica del Hospital Universitario de Salamanca. Dic. 1997 - Nov. 2007.

4. Discusión

El hepatocito se infecta con el virus de la hepatitis C y posteriormente se produce una nueva célula hepatocitaria al fusionarse la célula hepática senescente (24) con la célula madre hepática o con la célula precursora hepática (oval). (4), (23). La denominaremos: célula simbiooncomutagénica hepática, la cual luego de reproducirse aceleradamente por el influjo viral (22) y en un contexto anatomopatológico de fibrosis inflamatoria, que puede corresponder a hepatitis crónica o a cirrosis, la misma presenta escape apoptótico y al salvar la apoptosis se sigue reproduciendo gracias a que las células inmunes también presentan agotamiento celular. Al activarse las vías oncogénicas y del desarrollo tumoral respectivamente se produce una catástrofe simbiooncomutagénica. (16), (26), (28). Esta nueva célula bajo el constante influjo de marcadores de superficie y producción exagerada de telomerasa desembocan en la producción o génesis del hepatocarcinoma y células madre acompañantes que derivan en células con capacidad de desarrollo en tejidos aberrantes incluyendo vasos de neoformación para el crecimiento tumoral y más adelante el desarrollo de metástasis. Telomerasa y telómeros tumorales alcanzan la inmortalidad del hepatocarcinoma. (3), (33).

La telomerasa es la máquina del tiempo pero con dos túneles del campo unificado en un solo plano: la inmortalidad. El primer túnel es la vida como la conocemos. El segundo túnel es la neoplasia (muerte humana). Las células fetales, embrionarias y germinales, así como las neoplásicas expresan telomerasa. Recientes investigaciones muestran la importancia relevante entre telomerasa y cáncer, células madre, envejecimiento y síndromes hereditarios. (14). ¿Es una relación causa-efecto? Nuestra hipótesis es que sí. Primero la telomerasa y luego el tumor en nuestro modelo biomédico de hepatitis crónica y cirrosis hepática asociada a virus C (25) con posterior evolución a hepatocarcinoma. (15), (27). Probablemente la presencia de p53 mutado (41%) y la α fetoproteína (62%) sean consecuencia de mayor anaplasia de estos hepatocarcinomas (p: ns). (17), (19).

5. Conclusiones

1. Predominio de pacientes del sexo masculino en los tres grupos.
2. Recomendación de la prueba inmunohistoquímica p53, como factor de riesgo y diagnóstico en cirrosis hepática virus C y hepatocarcinoma, respectivamente.
3. Se recomienda ampliar el estudio con antígenos de superficie de la célula hepatocitaria tumoral para aclarar su origen jerárquico Bi ó multipartito por reacción en cadena de la polimerasa (P.C.R), según la teoría de la fusión celular en la simbiocomutagénesis del Hepatocarcinoma y VHC.

REFERENCIAS

- Allende Daniela S and Yerian Lisa. Immunohistochemical Markers in the Diagnosis of Hepatocellular Carcinoma. *Pathology Case Reviews* 2009; 14: 40-46.
- ARMSTRONG GL, ALTER MJ, MCQUILLAN GM, MARGOLIS HS. The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. *Hepatology*, 2000; 31: 777-782.
- BLACKBURN, EH. Switching and signaling at the telomere. *Cell*, 2001; 106: 661-673.
- Bosch Barrera J, Lopez – Picazo Gonzalez J.M., Garcia – Foncillas López J, Prósper Cardoso F. Células Madre y Cáncer: dilucidando el origen de la célula madre tumoral. *Rev Med Univ Navarra/Vol 51, N°2, 2007, 14 – 17.*
- CAPOCACCIA RICCARDO, Sant Milena et al. Hepatocellular carcinoma: trends of incidence and survival in Europe and the United States at the end of the 20th century. *AM.J. Gastroenterol*, 2007; 102: 1661-1670.
- CHOO Q-L, KUO G, WEINER AM, JONES DB, et al. Isolation of Ac DNA clone derived from A blood-borne non-A, non-B viral hepatitis genome. *Science*, 1989; 244: 359-362.
- DAVILA JA, MORGAN RO, SHAIB Y, MCGLYNN KA, EL-SERAG HB. Hepatitis C infection and the increasing incidence of hepatocellular carcinoma: a population-based study. *Gastroenterology*, 2004; 127: 1372-1380.
- EL-SERAG HB, DAVILA JA, PETERSEN NJ, MCGLYNN KA. The continuing increase in the incidence of hepatocellular carcinoma in the United States: an update. *Ann Intern Med*, 2003; 139: 817-823.
- EL-SERAG HB, MASON AC. Rising incidence of hepatocellular carcinoma in the United States. *N Engl J Med*, 1999; 340:745-750.
- EL-SERAG HB, MASON AC. Risk factors for the rising rates of primary liver cancer in the United States. *Arch Intern Med*, 2000; 160: 3227-3230.
- EL-SERAG HB, RUDOLPH KL. Hepatocellular carcinoma: Epidemiology and Molecular Carcinogenesis. *Gastroenterology*, 2007; 132: 2557-2576.
- EL-SERAG HB. Hepatocellular carcinoma: recent trends in the United States. *Gastroenterology*, 2004; 127 (5 suppl 1): S27-S34.
- FAUCI, Anthony S; BRAUNWALD, Eugene, et al. *Harrison. Principios de Medicina Interna*. McGraw Hill, 17^a Edición. 2008.
- FIBBE WILLEM, E. Telomerase mutatiois in aplastic anemia. Editorials. *N Engl J Med*, 2005; 352: 1481-1482
- FRÖHLING, Stefan; DÖHNER, Hartmut. Chromosomal abnormalities in cancer. *New England. J Med*, 2008 August 14; 359: 722-734.
- HAYFLICK, L. The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res*, 1965; 37: 614-636.
- LANED P. p53: guardian of the genome. *Nature*, 1992; 358: 1516.

LLOVET, et al. Presentación oral, 2007. Annual Meeting, 4 junio 2007. Chicago, IL, USA.

MITCHELL, Peter. Acoplamiento de la fosforilación a la transferencia de electrones e hidrógeno mediante un mecanismo de tipo quimiosmótico. *Nature*, 1961; Volume 191: 144-148.

NEAL KEITH, R. On Behalf of the Trent Hepatitis C Study Group. *GUT* 2007; 56: 1098-1104.

PARKIN DM, BRAY F, et al. Estimating the world cancer burden: Globocan 2000. *Int J Cancer* 2001; 94:153-156.

PAZ BOUZA, José Ignacio. Aspectos histopatológicos de diferentes órganos del ratón inoculados con virus herpes simples de tipo I y II. Tesis Doctoral. Facultad de Medicina. Universidad de Salamanca. España. 1978.

POTENZA, Leonardo; LUPPI, Mario; BAROZZI, Patrizia; et al. HHV- GA syncytial giant-cell hepatitis. *New England. J Med*, Number 6. 2008 August 7; 359: 593-602.

ROSENTHAL, Nadia. Prometheus's Vulture and stem-cell promise. Number 3. *New England. Journal of Medicine*, 2003 July 17; 349: 267-274.

RUDOLPH KL, CHANG S, MILLARD M, SCHREIBER-AGUS N, DEPINHO RA. Inhibition of experimental liver cirrhosis in mice by telomerase gene delivery. *Science*, 2000; 287: 1253-1258.

SÁNCHEZ POSADA, Raúl. Características biológicas de la célula en el cáncer de colon con metástasis ganglionares. Tesis Doctoral. Salamanca, 2007.

SUN WEI, XING BAOCAI et al. Proteome analysis of hepatocellular carcinoma by two-dimensional difference gel electrophoresis. *Molecular and Cellular proteomics*, 2007; 6: 1798-1808.

TANG, Weiliang; LAZARO, Catherine A; et al. Responses of non transformed human hepatocytes to conditional expression of full-length hepatitis C virus open reading frame. *Am J Pathol*, 2007; 171: 1831-1846.

WONG JB, MCQUILLAN GM, MCHUTCHISON JG, POYNARD T. Estimating future hepatitis C morbidity, mortality, and costs in the United States. *Am J Public Health*, 2000; 90: 1562-1569.

YOSHIZAWA H. Hepatocellular carcinoma associated with hepatitis C virus infection in Japan: projection to other countries in the foreseeable future. *Oncology*, 2002; 62 (suppl 1): 8-17.

YU MW, CHEN CJ. Elevated serum testosterone levels and risk of hepatocellular carcinoma. *Cancer Res*, 1993; 53: 790-794.

YU MW, YANG YC, YANG SY, et al. Hormonal markers and hepatitis B virus-related hepatocellular carcinoma risk: a nested case-control study among men. *J Natl Cancer Inst*, 2001; 93: 1644-1651.

ZEN YOH, FUJII TAKAHIKO et al. Histological and culture studies respect to ABC62 expression support the existence of a cancer cell hierarchy in human hepatocellular carcinoma. *AJP*, 2007; 170: 1750-1762

***Nota de origen:**

*Artículo derivado de la Tesis Doctoral: "Estudio inmunohistoquímico de las lesiones hepáticas producidas por el virus de la hepatitis C y su relación con el hepatocarcinoma". Del DR. Martín Varela Vindas Ph. D. Director de la tesis: Dr. José Ignacio Paz Bouza Ph. D. Departamento de Biología Celular y Patología, Universidad de Salamanca, España. Calificación: Sobresaliente Cum Laude.
docmartinvarela@yahoo.com

La Rugoscopia Palatina Humana como Factor Fundamental de Identificación Personal

Noralia Ramírez Chávez¹, Alfonso Sotelo y Soto Gustavo², Jorge de Jesús Serena Melo³, Guillermo Hernández Lira⁴, Rafael Castillo Pratz⁵, y Rafael Alonso Nuñez⁶

¹ Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. nramirez@uv.mx

² Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. gsotelo@uv.mx

³ Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. jserena@uv.mx

⁴ Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. guillhernandez@uv.mx

⁵ Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. cpratz@uv.mx

⁶ Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. raalonso@uv.mx

1. Introducción

Este trabajo es factible de entender, aplicar y ubicar en un contexto universal, el caso que nos ocupa se sitúa para los fines prácticos de investigación antropométrica, en la región central del estado de Veracruz, y con circunscripción a la zona de Xalapa.

Al decir que su contextualización pudiese ser universal, nos estamos refiriendo a que la rugoscopia palatina es común a todos los seres humanos, conformada por las eminencias o crestas formadas en el tercio anterior de la mucosa palatina y que distribuidas a cada lado de la línea media o rafe palatino, dan a esta zona un aspecto similar a las nervaduras de una hoja de planta.

Las variaciones en características anatómicas tan importantes como las rugas palatinas, que vuelven al hombre individual e irrepetible, no obstante la cantidad de seres humanos que poblamos el mundo, debe tener una explicación y es necesario que se investigue de manera mas profunda y cuidadosa. Este modesto trabajo es el principio de una línea de investigación que pretende ser desarrollada por los académicos de la Facultad de Odontología, Xalapa. Deseamos también favorecer la divulgación para que esta parte de identificación personal, sea considerada como parte integral de la ficha de identificación personal en las historias clínicas odontológicas.

2. Antecedentes

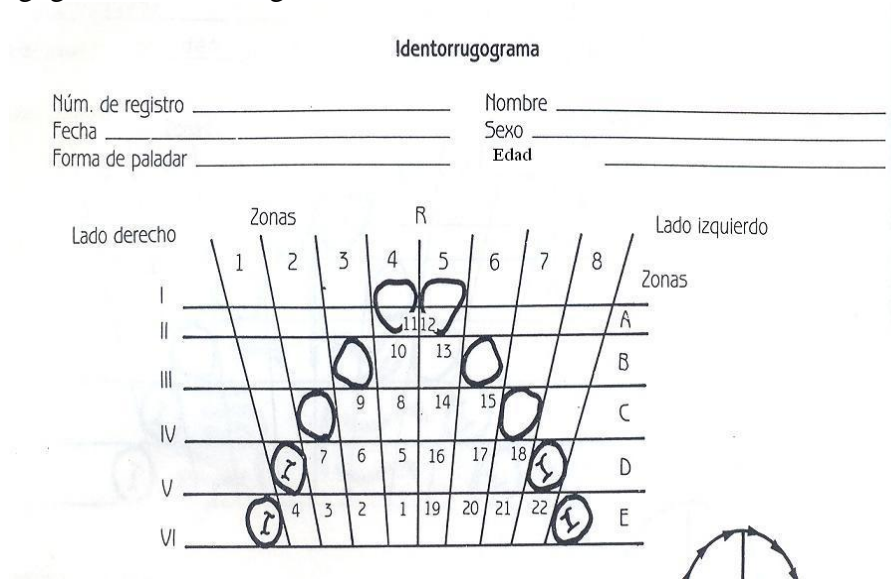
Algunos investigadores estudiaron la presencia de las rugas palatinas en diversos años, sin embargo hasta 1938 fueron minuciosamente estudiadas por el Dr. Luis Silva, quien las homologó en cierto sentido con las impresiones dactilares, dado que: a) son invariables en su forma; b) son permanentes a lo largo de la vida; c) son absolutamente diferentes de un individuo a otro; y, d) son absolutamente diferentes entre si. Este investigador se ocupó extensamente de su empleo en la identificación individual, aportando un sistema para describirlas y registrarlas con dicho propósito, poco después en 1955 el Dr. Carrea, aportó nuevos estudios post/natales. Generalmente se han empleado para identificación de uso legal, pero en el caso que ahora nos ocupa la perspectiva de aplicación es genética.

3. Hipótesis de trabajo

Alguna de las rugas del total que conforma la fórmula rugoscópica de cada individuo corresponde en particular a una presencia en las fórmulas rugoscópicas de una familia, ya que se asume tienen una carga genética. La hipótesis a comprobar es la presencia de algún tipo y posición de ruga como factor de marcaje genético hereditario.

4. Metodología

Es una investigación básica aplicada, el objeto de estudio se ubica en ocho familias las cuales cuentan con un mínimo de tres generaciones y un máximo de cuatro, se realiza el estudio con el objetivo de elaborar una base de datos de sus rugoestenografías o fórmulas rugoscópicas, que es el gráfico en el que se emplean números y letras convencionales para describir el tipo de rugas palatinas, así como su posición en cada individuo. Complemento muy importante es que se debe tomar una fotogrugoscopia sobre los modelos de yeso. La población estudiada son ocho familias con un total de 44 casos. A los cuales se les elabora un identorrugograma como el siguiente:



Cuadro de resultados de comparación y coincidencias de las rugas palatinas de los miembros de cada familia.

Familia 1. Responsable: alumno Néstor Martínez Fuentes

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|----------------|-------------------|---|
| 1 | F1.1 | Abuelo paterno | 5597 | * Segunda ruga derecha, posición canino y forma curva(2) |
| 2 | F1.2 | Abuelo materno | 5598 | |
| 3 | F1.3 | Padre | 5599 | * Segunda ruga derecha, posición canino y forma curva(2) |
| 4 | F1.4 | Madre | 5600 | |
| 5 | F1.5 | Hija 1 | 5601 | |
| 6 | F1.6 | Hija 2 | 5602 | * Segunda ruga derecha, posición canino y forma curva (2) |

Familia 2. Responsable: alumno Alonso Martínez Ochoa

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|----------------|-------------------|--|
| 7 | F2.1 | Abuelo materno | 5603 | |
| 8 | F2.2 | Padre | 5604 | *Cuarta ruga izquierda, posición de primer premolar y forma sinuosa (4) |
| 9 | F2.3 | Hijo 1 | 5607 | * Cuarta ruga izquierda, posición de primer premolar y forma sinuosa (4) |
| 10 | F2.4 | Hijo 2 | 5608 | * Cuarta ruga izquierda, posición de primer premolar y forma sinuosa (4) |
| 11 | F2.5 | Madre | 5609 | |

Familia 3. Responsable alumna: Sindi Vásquez libreros

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|----------------|-------------------|---|
| 12 | F 3.1 | Abuelo paterno | 5610 | |
| 13 | F 3.2 | Padre | 5611 | * Primera ruga izquierda posición entre lateral y canino, forma curva hacia abajo (2) |
| 14 | F 3.3 | Madre | 5612 | |
| 15 | F 3.4 | Hija | 5613 | * Primera ruga izquierda posición entre lateral y canino, forma curva hacia abajo (2) |
| 16 | F 3.5 | Nieta | 5614 | |

Familia 4 Responsable: alumna Adeli Karina Trakonis Herrera

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|----------------|-------------------|--|
| 17 | F4.1 | Abuela paterna | 5615 | *Segunda ruga izquierda en forma curva orientada hacia arriba y posición canino (2) |
| 18 | F4.2 | Padre | 5616 | * Segunda ruga izquierda en forma curva orientada hacia arriba y posición canino (2) |
| 19 | F4.3 | Madre | 5617 | |
| 20 | F4.4 | Hija | 5618 | * Segunda ruga izquierda en forma curva orientada hacia arriba y posición canino (2) |

Familia 5 Responsable: alumna Dannay Croda Alarcón

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|----------------|-------------------|--|
| 21 | F5.1 | Abuela materna | 5812 | |
| 22 | F5.2 | Abuelo materno | 5811 | *Segunda ruga derecha en forma recta, situada en zona de premolares (1) |
| 23 | F5.3 | Madre | 5809 | * Segunda ruga derecha en forma recta, situada en zona de premolares (1) |

Familia 6 Responsable: alumna Irais Vargas Jiménez

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|----------------|-------------------|--|
| 25 | F6.1 | Hija | 5820 | + Primera ruga izquierda curva hacia abajo posición canino (2) |
| 26 | F6.2 | Madre | 5821 | + Primera ruga izquierda curva hacia abajo posición canino (2) |
| 27 | F6.3 | Padre | 5822 | * Primera ruga derecha curva hacia abajo y posición entre canino y premolar (2) ~ Segunda ruga izquierda forma recta, posición entre canino y premolar. |
| 28 | F6.4 | Abuela materna | 5823 | |
| 29 | F6.5 | Abuelo materno | 5824 | + Primera ruga izquierda curva hacia abajo posición canino (2) |
| 30 | F6.6 | Abuela paterna | 5825 | *Primera ruga derecha curva hacia abajo y posición entre canino y premolar (2) |
| 31 | F6.7 | Abuelo paterno | 5826 | ~ Segunda ruga izquierda forma recta, posición entre canino y premolar. |

Familia 7. Responsable C. D. Seth Daniel Contreras Báez

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|-------------------|-------------------|---|
| 32 | F7.2 | Madre | 5828 | |
| 33 | F7.3 | Hija | 5829 | * Primera ruga derecha forma angular acostada posición de canino (3). |
| 34 | F7.4 | Abuela materna | 5830 | |
| 35 | F7.5 | Bisabuela materna | 5831 | |
| 36 | F7.6 | Abuela paterna | 5832 | |
| 37 | F7.7 | Padre | 5833 | * Primera ruga derecha forma angular acostada posición de canino (3). |
| 38 | F7.8 | Abuelo materno | 5834 | |
| 39 | F7.9 | Abuelo paterno | 5835 | |

Familia 8 Responsable alumna: Aracely Vásquez Galicia

| No. progr | No. de modelo | Parentesco | Folio fotográfico | Observaciones |
|-----------|---------------|----------------|-------------------|--|
| 40 | F8.1 | Abuelo materno | 5856 | |
| 41 | F8.2 | Abuela materna | 5857 | *Primera ruga izquierda posición canino en forma Ye (6) |
| 42 | F8.3 | Madre | 5858 | * Primera ruga izquierda posición canino en forma Ye (6) |
| 43 | F8.4 | Padre | 5859 | |
| 44 | F8.5 | Hijo | 5860 | * Primera ruga izquierda posición canino en forma Ye (6) |

5. Resultados

La frecuencia esperada es favorable, pues existe una coincidencia en todas (100%) las familias. Se establece las coincidencias que son favorables en número de 17, para el total de casos que son cuarenta y cuatro.

| Variable | Ss | Total de casos | Número de Familias | Tipos de rugas |
|-------------------|----|----------------|--------------------|----------------|
| Bisabuela materna | 1 | 44 | 8 | 2 |
| Abuelo paterno | 4 | 44 | 8 | 3 |
| Abuela paterna | 3 | 44 | 8 | 1 |
| Abuelo materno | 6 | 44 | 8 | 4 |
| Abuela materna | 4 | 44 | 8 | 6 |
| Madre | 8 | 44 | 8 | |
| Padre | 7 | 44 | 8 | |
| Hijo | 10 | 44 | 8 | |
| Nieta | 1 | 44 | 8 | |

| Media | Mediana | Moda | Frecuencia | Valor mínimo | Valor máximo | Desv Standar |
|----------|---------|------|------------|--------------|--------------|--------------|
| 3.789833 | 4 | 1 | 2 | 1 | 10 | 3.100179206 |
| | | | | | | |

REFERENCIAS

- Ciocca, G. L. Diversos sistemas de identificación odontológica por medio de aparatos protésicos, por Rugoscopía .. 11 de enero del 2007.
- Correa, R.A, (1990) “*Estomatología Forense*” Editorial Trillas, México
- Medrano, M. J., Pérez F.(2005) “*Expediente Clínico Odontológico*”.Edit. Trillas. 1ª. Edición. México D. F.
- Rojas, N; (1950).”*Medicina Legal*” Edit.El Ateneo, -
- Quintero, L.C. En “*Odontología identificación positiva por prótesis total.*” Reporte de caso. No. 1, 2006

Aportaciones a la Medida de la Fiabilidad del Instrumento Purpose in Life (PIL)

Dr. José David Urchaga Litago¹, Lic. Eliseudo Salvino Gomes², y Dr. Antonio Sánchez Cabaco³

¹Universidad Pontificia de Salamanca. España. urchaga@upsa.es

²Universidad Pontificia de Salamanca. España. salvinno@uol.com.br

³Universidad Pontificia de Salamanca. España. asanchezca@upsa.es

1. Resumen

El test PIL (Purpose In Life) es uno de los instrumentos más importantes para evaluar el sentido de la vida de las personas. Todas las investigaciones lo describen como muy fiable. Presentamos un estudio transcultural (preuniversitarios de Brasil y España) que cuestiona la fiabilidad de dicho instrumento para la adaptación española. Además, el programa informático SPSS no detecta en un primer momento dicho error en los ítems “invertidos”, por lo que proponemos un estudio gráfico de la distribución de los ítems como paso previo al cálculo de la fiabilidad. En el estudio participaron 621 estudiantes de Brasil (363 mujeres, 258 hombres) y otros tantos de España (351 mujeres, 270 hombres), con una edad media de 16,7 años.

2. El instrumento de medida

El PIL fue elaborado en los Estados Unidos por Crumbaugh y Maholick (1969). La prueba consta de tres partes, pero en este trabajo sólo nos referiremos a la primera, ya que las otras no corresponden a preguntas de respuesta cerrada. La primera parte consta de 20 ítems donde cada sujeto se posiciona en una escala tipo likert (1-7) entre dos sentimientos extremos. Por ejemplo, el ítem 1:

| | | | | | | | | |
|------------------------------|---|---|---|---|---|---|---|--------------|
| - <i>Generalmente estoy:</i> | | | | | | | | |
| Totalmente aburrido | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Entusiasmado |

La escala consta de 10 ítems de corrección directa (a mayor puntuación mayor sentido de la vida), y otros diez de corrección inversa, ya que los dos sentimientos están encontrados del más positivo al negativo. Por ejemplo, el ítem 7:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|-----------------------------|
| - <i>Después de jubilarme, me gustaría:</i> | | | | | | | | |
| Hacer cosas que me han interesado | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Vaguear el resto de mi vida |

Una vez corregidas las puntuaciones en los ítems inversos, se suman todas las puntuaciones obteniéndose una única puntuación total que se valora en función de diferentes baremos publicados.

3. La fiabilidad psicométrica

La fiabilidad en psicometría se entiende como precisión y estabilidad de las puntuaciones, por lo que hace referencia a la exactitud y falta de error en la medida (Fuentes, 2001; Martínez, 1996; Morales y otros, 2003). En la medida de constructos está muy ligada a las técnicas estadísticas multivariantes, ya que no es posible obtener mediciones directas, es decir, no existen “cuestionarios” que con una sola pregunta permitan medir dichos constructos. Por eso, es necesario construir tests con un número de preguntas que midan parte de dichos constructos. Si el test fuera fiable, tendría que suceder que si midiéramos diferentes veces, y el sujeto no hubiera variado en dicha dimensión, se deberían obtener las mismas puntuaciones. Para ello es necesario que cada una de las preguntas del test fuera estable en las diferentes mediciones. Hay varios procedimientos para estimar la fiabilidad de un test, pero los dos más utilizados son el procedimiento de alfa de Cronbach (que se puede interpretar como una aproximación a una hipotética correlación media entre todos los ítems del test) y el de dos mitades (correlación entre una mitad aleatoria de los ítems con la otra mitad). Este último procedimiento necesita de una corrección, ya que el resultado inicial sería sobre un test con la mitad de los ítems originales, y se sabe que la fiabilidad aumenta con el número de ítems (la fórmula correctora más utilizada para estimar la fiabilidad final es la propuesta por Spearman-Brown).

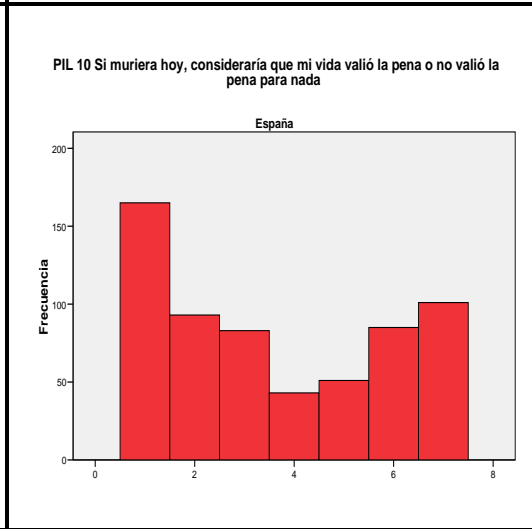
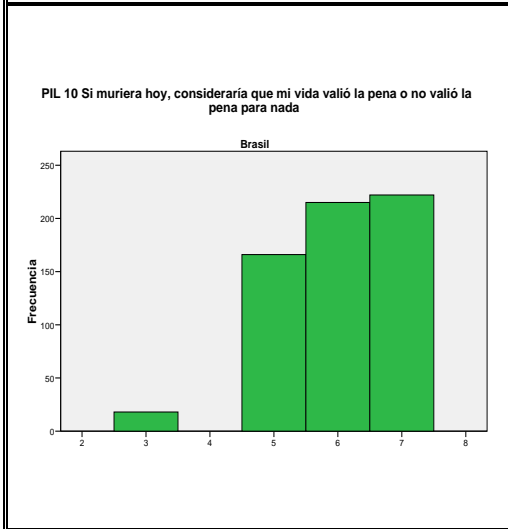
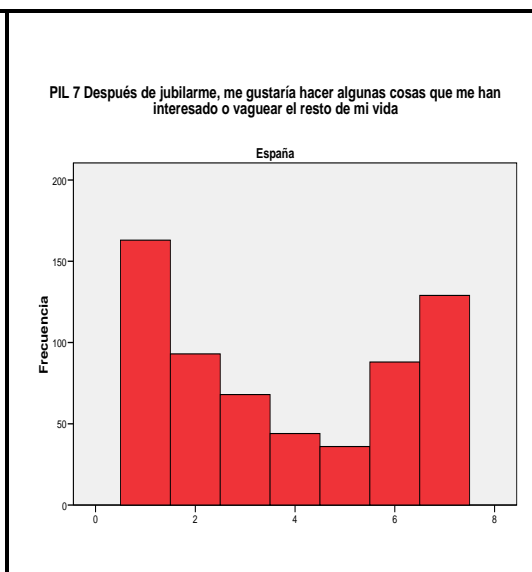
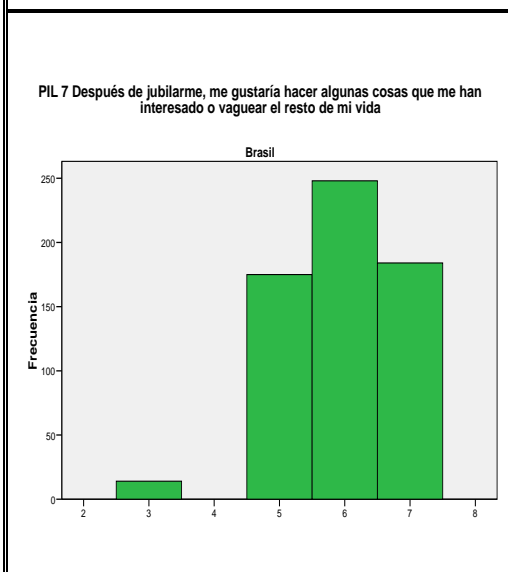
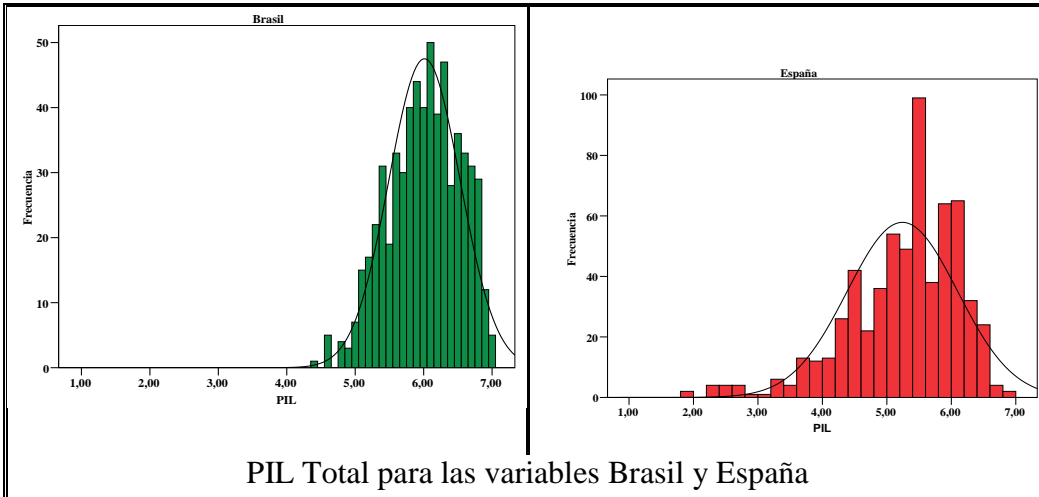
Así, en nuestro caso, de cada sujeto tenemos 21 puntuaciones (una por cada pregunta del test, más la puntuación total).

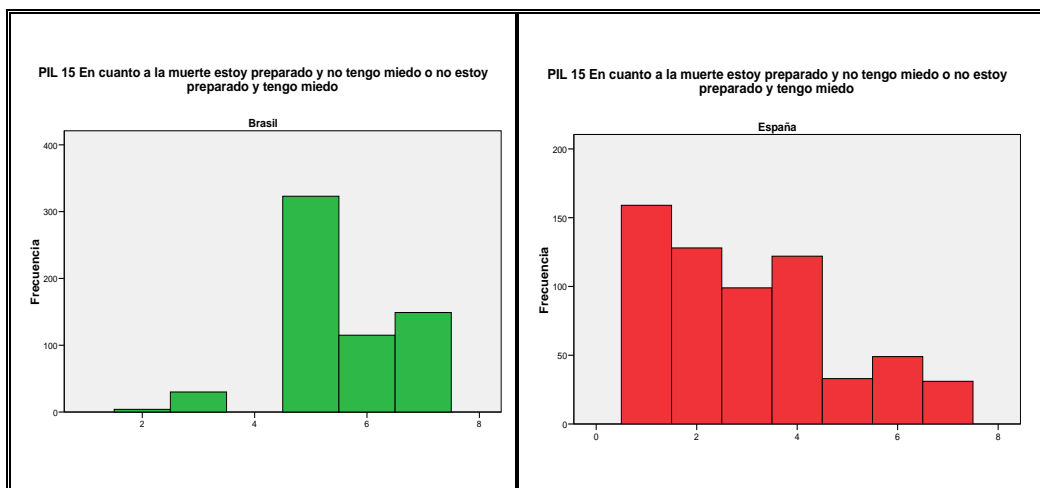
4. Los resultados obtenidos

Para los cálculos nos servimos del programa SPSS. Los resultados mostraban, en un primer momento, una gran fiabilidad para el test PIL, pues el índice de *alfa de Cronbach* era de 0,911, el cual es un buen resultado en comparación con otros estudios: Shek (1986) en China (0.84 para el alfa de Cronbach y 0.82 para el coeficiente de Guttman); Reker y Cousins (1979) en Estados Unidos (0.87 para Spearman-Brown y 0.79 para el procedimiento test-retest); Teixeira et al (2006) en Portugal y Brasil (0.70 para el alfa de Cronbach); Noblejas (1994, 2000) con una muestra española obtuvo un alfa de 0,88; Sánchez Cabaco, Urchaga y otros (2006), con universitarios españoles obtuvieron un alfa de 0,82.

El procedimiento de *dos mitades*, nos ofrecía una muy buena fiabilidad (0,89 es la correlación entre las dos mitades; 0,94 si se realiza la corrección de Spearman-Brown).

El problema surgió al estudiar la distribución de los ítems de forma individualizada y encontrarnos que algunos (en especial los que se plantean de forma inversa) distribuían las puntuaciones de forma diferente a la distribución de las puntuaciones totales del test. A continuación presentamos algunas de estas distribuciones, diferenciando si son de la muestra brasileña o española:





Como se puede observar, en la muestra española, a diferencia de la brasileña, las distribuciones son muy diferentes en dichos ítems en comparación con la distribución de las puntuaciones totales del test. Lo lógico es que todos los ítems, si la escala es muy fiable, mostrarán similares formas en sus distribuciones a la de la de la distribución de las puntuaciones totales en el test.

Esto nos lleva a cuestionarnos la fiabilidad psicométrica que pudieran presentar los ítems que mostraban distribuciones tan diferentes. Planteamos la posibilidad de que parte de la muestra española no hubiera respondido correctamente a gran parte de los ítems “invertidos” (ya fuera por una mala redacción, por dificultades de comprensión, o por cualquier otro motivo).

Se recalculó la fiabilidad por el procedimiento de dos mitades, pero introduciendo las variables en la opción de SPSS de tal forma que tomara como una mitad los ítems directos, y la otra los “inversos” (el SPSS compara los 10 primeros ítems introducidos en su cuadro de diálogo con los diez últimos). Fuera de lo esperado, se obtuvo una fiabilidad paradójicamente baja: 0,085 para el caso español y 0,529 para el caso brasileño.

| | | Brasil | España |
|-------------------------------|-------------|--------|--------|
| Alfa de Cronbach | Parte 1 (a) | 0,396 | 0,833 |
| | Parte 2 (b) | 0,753 | 0,800 |
| Correlación entre formas | | 0,359 | 0,085 |
| Coeficiente de Spearman-Brown | | 0,529 | 0,157 |

a Los ítems son: 1, 2, 5, 7, 10, 14, 15, 17, 18, 19.

b Los ítems son: 3, 4, 6, 8, 9, 11, 12, 13, 16, 20.

5. Conclusiones

Lo paradójico está en que cada parte obtiene altos valores, y que el alfa de Cronbach obtenido con la escala total sea elevado, pero que al correlacionar ambas partes (comparando los ítems directos y los invertidos) la correlación sea tan baja. Esto implica que el procedimiento de cálculo de la fiabilidad por los procedimientos de dos mitades y

alfa de Cronbach pueden llevar a conclusiones muy erróneas según el cómo se agrupan los ítems en el procedimiento de dos mitades. No conocemos ninguna publicación –al menos en español- que haga referencia a este posible error. Recomendamos un estudio visual de las distribuciones de los ítems (incluida la distribución de la puntuación final del test), y si hubiera algún ítem dudoso, realizar algún estudio que pudiera mostrar si esas diferencias son significativas o no (por ejemplo con las pruebas de Levene, la de Kolmogorov-Smirnov, la de Friedam o la W de Kendall).

REFERENCIAS

- Crumbaugh, J. C. y Maholick, L. T. (1969). *Manual of instructions for the Purpose in Life Test*. Saratoga: Viktor Frankl Institute of Logotherapy.
- Teixeira, M. L.; DeDomenico, S. R.; Mello, N. O.; Almeida, C. M. A.; & Almeida, F. J. R. (2006). Sentidos de vida: um estudo transcultural entre gestores brasileiros e portugueses. *Organizações & Sociedade*, 13, (38), 15-30.
- Fuentes Blanco, J.M^a (2001). *Fundamentos de psicometría*. Salamanca: Amarú.
- Martínez Arias, R. (1996). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Morales Vallejo, P., Urosa Sanz, B. y Blanco Blanco, A. (2003). *Construcción de escalas de actitudes tipo Likert*. Salamanca: Hespérides.
- Noblejas, M. A. (1994). *Logoterapia. Fundamentos, principios y aplicación. Una experiencia de evaluación del logro interior de sentido*. Tesis Doctoral, Universidad Complutense de Madrid.
- Noblejas, M. A. (2000). Fiabilidad de los tests PIL y Logotest. *NOUS: Boletín de Logoterapia y Análisis Existencial*, (4), 81-90.
- Reker, G.T. y Cousins, J.B. (1979). Reliability of the Seeking Of Noetic Goals (SONG) and Purpose In Life (PIL) tests. *Journal of Clinical Psychology*. 35, (1), 85-91.
- Sánchez Cabaco, A., Urchaga Litago, J.D. y otros (2006) (documento no publicado). *Evaluación de competencias actitudinales (sentido de la vida) en universitarios salmantinos: implicaciones para la adaptación al EEES*. Salamanca: Universidad Pontificia de Salamanca, Memoria de proyecto de investigación.
- Shek, D.T.L. (1986). The purpose in life questionnaire in a chinese context: some psychometric and normative data. *Chinese Journal of Psychology*. 28, (1), 51-60.

Sistema para Evaluar el Grado de Percepción de los Clientes sobre una Organización Prestadora de Servicios

Juan de Dios Villaseca Galván

Sub-Dirección de Enseñanza e Investigación. Hospital Regional de Alta Especialidad “Dr. Juan Graham Casasús”. Villahermosa, Tabasco. México. pae_2002@hotmail.com

1. Introducción

Philip Kotler propone cuatro factores básicos que definen la predilección del cliente por un producto o servicio, siendo éstas: a) Producto, b) Servicio, c) Imagen de la organización e d) Imagen que genera el personal. El objetivo del presente trabajo fue crear un sistema computarizado basado en la plataforma de Excel que permita otorgar una puntuación a los factores que según Kotler determinan la decisión de compra o servicio del cliente.

2. Materiales y Métodos

Se consideró solo tres factores: Servicio (S), Imagen de la organizacional (I) e Imagen que genera el personal (P). Se eliminó producto

Cada factor anterior se descompuso en sub-factores que conforman el concepto integral en sí. Una vez escogidos los subfactores, se procedió a enlistar cada uno de sus antónimos, ya que se buscó establecer un diferencial semántico (método de Osgood). La escala de calificación aplicada para el diferencial semántico fue la de Likert, con un rango de cinco puntos entre cada adjetivo calificativo y su antónimo.

Diseño del formato de la encuesta: La encuesta se compone de treinta items agrupados en los tres factores seleccionados S, I, P.

Determinación del universo (N): Se realizó una análisis de la población que componen el universo hospitalario quedando conformado por:

- a) Cliente Interno (C.I.) en cual se integra su vez por: Empleado Federal, Empleado Estatal, Personal de Contrato, Estudiantes de Servicio Social y Estudiantes de Prácticas Profesionales.
- b) Cliente Externo (C.E.) se integra por: Pacientes de Consulta Externa, Pacientes Hospitalizados, Familiar de Pacientes de Consulta Externa, Familiar de Pacientes hospitalizados y Proveedores (éstos últimos se consideran clientes una vez que esperan su pago en tiempo y forma).

En el caso de los C.I. se debió realizar un análisis de la estructura organizacional en todos los turnos en los que funciona la institución.

Muestreo: Se aplicó un muestreo Estratificado por afijación proporcional para poder determinar el número de encuestados a los que se debía aplicar la encuesta, debiéndose considerar la siguiente formula para poblaciones finitas de menos de 100,000 habitantes la cual se establece de la siguiente manera:

$$n = \frac{z(N)(p)(q)}{E^2(N-1) + [z(p)(q)]}$$

Y para la cual había que determinar el error de estimación, siendo calculado como sigue:

$$E = \frac{\sqrt{z(p)(q)}}{n}$$

Una vez establecido el tamaño de la muestra poblacional (n), se calculó el porcentaje de la muestra en el universo poblacional (N) por medio de una regla de tres $N = 100\%$ con lo cual se obtiene el porcentaje de la cuota a aplicar en cada una de las $n = x\%$ áreas hospitalarias y en cada uno de los turnos en lo que trabaja la institución, no importando el puesto o la jerarquía del C.I.

Cabe señalar que para el diseño del muestreo se consideró que el mercado meta es en sí una población con distribución normal y que por el mismo grado de confianza elegido (ya que se tenía una idea muy clara por la prueba piloto) se escogió el valor Z de 3.84 para dos sigmas con un grado de confianza de 95%.

Aplicación de la encuesta: La encuesta se aplicó al porcentaje obtenido del universo (N) entre la muestra poblacional calculada (n) para cada una de las distintas áreas que componen al hospital en cada uno de los turnos.

Sistema de recolección y captura: Las encuestas capturadas fueron procesadas a más tardar a la hora de su aplicación, asegurando con esto un método rápido y eficiente para su revisión y captura en Excel. Anotando en la esquina derecha el número de folio de la encuesta para su identificación, así como la fecha y hora de aplicación y captura, y la rubrica del capturista. Posteriormente el administrador de la información procede a verificar mediante muestreo al azar la captura correcta en la hoja de Excel programada para el procesamiento.

Como se mencionó anteriormente, la captura se realiza en una hoja de Excel programada con las formulaciones pertinentes de cada uno de los subfactores de S. I. y P. siendo calificados del 20 al 100, considerándose el 20 como el puntaje más bajo, el cual representa al adjetivo calificativo negativo. Logrando con esto obtener puntajes totales de cada uno de los subfactores o puntajes promedios de cada uno de los encuestados.

Calificación: La calificación se obtiene del promedio de cada uno de los subfactores, que componen los factores S. I. P. y que a su vez generan un promedio de cada factor, permitiendo comparar la calificación de percepción tanto de los C.I. como de los C.E. para percibir la diferencia de percepción.

3. Resultados

Los resultados obtenidos de la presente encuesta indican las siguientes inferencias según las tablas de resultados de cada uno de los factores siguientes:

Puntuación total final:

La escala de evaluación tiene la capacidad de general una puntuación general final, considerando el error de estimación que se genera en el muestreo, lo cual le puede otorgar cierta confiabilidad. La puntuación obtenida al final fue de 86.20 de un total de 100 punto, lo que establece a la institución con un puntaje aceptable en la percepción positiva que el

cliente en general tiene de ella (tanto el C.I. como el C.E.), con un error de estimación (E) de 0.04 resultando un intervalo de confianza de 82.83 en el rango inferior y de 89.74 en el rango superior. Por los resultado anteriores, se debe de considerar que el puntaje final no es suficiente, ya que los gastos de operación de una institución pública como esta, son desproporcionados para la percepción que se tiene de su servicio, lo que se podría mejorar sí se realizara estrategias de comunicación efectivas, ya que según los datos que proporcionan las tablas 1, 2 y 3 en su análisis de los subfactores, indican una probable falta de información, o en el mejor de los casos, un sistema de información deficiente.

Escala Servicio (S)

Se obtuvo un puntaje promedio entre el C.I. y el C.E. de 84.35 (de 100 puntos) lo que le ofrece un puntaje razonable en la percepción general de la escala, observándose una discrepancia de 0.02 dentro de los mismos 100 puntos, lo que da un índice bastante aceptable de congruencia en la percepción. Reconociéndose a un C.E. bastante conocer en el servicio que se brinda (Tabla 1). Sin embargo existe una discrepancia constante (aunque mínima) entre los C.I. y los C.E. atribuyéndose a un posible estado de carencia e insuficiencia de éstos últimos, quedando la posibilidad de que lo anterior sea analizado y estudiado en una investigación de campo a futuro.

Tabla 1

| S E R V I C I O | | | | |
|----------------------------------|--------------------|--------------------|--------------|-----------------------------------|
| ESCALAS | C.I. x̄ | C.E. x̄ | Disc. | PROMEDIO (C.I. + C.E.) |
| EFICIENCIA | 78.04 | 85.37 | 7.33 | 81.70 |
| OPORTUNIDAD | 76.54 | 75.18 | -1.36 | 75.86 |
| EXPERIENCIA | 92.15 | 90.34 | -1.81 | 91.25 |
| COSTO | 73.74 | 71.54 | -2.20 | 72.64 |
| UTILIDAD | 95.33 | 91.90 | -3.42 | 93.62 |
| CONFIABILIDAD | 92.71 | 90.71 | -2.00 | 91.71 |
| HONESTIDAD | 87.76 | 92.17 | 4.41 | 89.96 |
| HUMANISMO | 87.20 | 91.50 | 4.31 | 89.35 |
| ADECUACION | 88.50 | 91.69 | 3.18 | 90.10 |
| ABASTO | 71.78 | 63.11 | -8.67 | 67.44 |
| CALIFICACION PROMEDIO | 84.37 | 84.35 | -0.02 | 84.36 |

Escala Imagen Organizacional (I)

En esta escala el C.E. le otorga un puntaje ligeramente mayor en 1.01 % al C.I. ya que al parecer la imagen que la institución genera es positiva, sin embargo no es sobresaliente, ya que el puntaje en general es de 88.13 de 100 puntos, no logrando alcanzar los 90 puntos, sobre todo sí se toma en cuenta el alto costo de inversión que conlleva el mantener funcional a una institución como lo es el Hospital Regional Dr. Juan Graham Casasús. En los subfactores de Costo, al parecer el C.E. desconoce el verdadero costo que implica la prestación del servicio médico, así como la modernidad de sus instalaciones, equipo médico, paramédico y administrativo. Por último existe un puntaje de discrepancia importante en el factor calidad, ya que al parecer el C.E. desconoce los procesos de calidad por los que la institución ha estado pasando.

Tabla 2

IMAGEN ORGANIZACIONAL

| <u>ESCALAS</u> | <u>C.I.</u> <u>\bar{x}</u> | <u>C.E.</u> <u>\bar{x}</u> | <u>Disc.</u> | <u>PROMEDIO</u> <u>(C.I. + C.E.)</u> |
|----------------------------------|--|--|--------------|---|
| HUMANISMO | 89.72 | 94.88 | 5.16 | 92.30 |
| ORDEN | 85.42 | 90.81 | 5.39 | 88.11 |
| LIMPIEZA | 92.34 | 94.20 | 1.86 | 93.27 |
| COSTO | 76.07 | 74.33 | -1.74 | 75.20 |
| MODERNIDAD | 86.92 | 80.93 | -5.98 | 83.92 |
| COMODIDAD | 87.38 | 87.93 | 0.55 | 87.66 |
| EFICIENCIA | 86.21 | 92.60 | 6.39 | 89.41 |
| CALIDAD | 89.72 | 82.24 | -7.48 | 85.98 |
| DISCIPLINA | 90.00 | 95.20 | 5.20 | 92.60 |
| ADAPTACION | 92.43 | 93.20 | 0.77 | 92.82 |
| CALIFICACION PROMEDIO | 87.62 | 88.63 | 1.01 | 88.13 |

Escala Imagen del Personal (P)

Esta escala deja muy claro que la imagen que el personal de la institución genera al C.E. es positiva, ya que por lo visto la discrepancia a favor del C.I. es alta, atorgándole el C.I. un puntaje que tiende a ser muy superior (7.19 puntos de discrepancia de 100) siendo sobresalientes las subescalas de Amabilidad, Paciencia, Credibilidad, Ansiedad y Anímico (estado de ánimo).

Es de notarse también que el puntaje en general de percepción de esta escala no es tan alto como lo que se podría esperar, obteniendo un total promedio de 86.37 puntos, siendo mermado por los puntajes otorgados por los mismos C.I. ya que al parecer conocen sus deficiencias de manera puntual.

Tabla 3

IMAGEN DEL PERSONAL

| ESCALAS | C.I. \bar{X} | C.E. \bar{X} | Disc. | PROMEDIO (C.I. + C.E.) |
|----------------------------------|--------------------------------------|--------------------------------------|--------------|-----------------------------------|
| AMABILIDAD | 78.22 | 89.68 | 11.45 | 83.95 |
| PACIENCIA | 77.20 | 88.76 | 11.57 | 82.98 |
| CREDIBILIDAD | 81.40 | 90.77 | 9.37 | 86.09 |
| ANSIEDAD | 78.32 | 90.16 | 11.84 | 84.24 |
| DISCIPLINA | 84.02 | 89.49 | 5.47 | 86.76 |
| ANIMICO | 80.65 | 88.73 | 8.07 | 84.69 |
| RESPECTO | 87.57 | 91.84 | 4.27 | 89.70 |
| HUMANIDAD | 84.49 | 91.24 | 6.75 | 87.86 |
| EXPERIENCIA | 89.16 | 91.45 | 2.30 | 90.31 |
| DISPOSICION | 86.73 | 87.53 | 0.80 | 87.13 |
| CALIFICACION PROMEDIO | 82.78 | 89.97 | 7.19 | 86.37 |

REFERENCIAS

Kotler, Philip; Armstrong, Gary. Mercadotecnia. Prentice-Hall Hispanoamericana. México, DF, MX. 1994

Carmen María Salvador Ferrer. La percepción del cliente de los elementos determinantes de la calidad del servicio universitario, características del servicio y habilidades profesionales. Universidad de Almería.

William J. Stevenson. Estadísticas para administración y economía. Harla. Harper & Row Latinoamericana. México, D.F. 1981

Salvador Mercado. Investigación de mercados –Teoría y práctica. Editorial Pac, S.A. de C.V. México D.F. abril 1999

Andrés Mazaira Castro, Ana Dopico Parada, Encarnación González Vázquez. Incidencia del grado de orientación al mercado de las organizaciones empresariales en el desarrollo de las capacidades estratégicas del marketing. Universidad de Vigo

Modelos Matemáticos para la Estimación del Desarrollo de Novillas: Edad y Peso a Primera Concepción

Roberto Castillo Tlapa¹, Belisario Domínguez Mancera², Antonio Hernández Beltrán³, José Manuel Martínez Hernández⁴, Patricia Cervantes Acosta⁵, Blanca Catalina Colin Ibarra⁶

Facultad de Medicina Veterinaria y Zootecnia. Universidad Veracruzana. ¹rcastillo@uv.mx, ²beldominguez@uv.mx, ³posfmvz@gmail.com, ⁴manuelmartinez@uv.mx, ⁵biomol@gmail.com, ⁶bcolini@hotmail.com

Introducción

En investigaciones con animales para determinar ciertos efectos, es indispensable que se tomen varias observaciones sobre el mismo animal en una secuencia de períodos de tiempo sucesivos, esta condición se conoce como modelo experimental basado en "mediciones repetidas"; esta metodología es usada cuando se determinan las curvas de crecimiento (Altamirano *et al.*, 2006). El análisis para este tipo de datos requiere una atención especial, ya que establecen patrones de correlación entre las observaciones que pueden arrojar conclusiones inválidas al evaluarlas con los procedimientos utilizados para otros modelos que no involucran el factor tiempo y pudieran representar un desperdicio de información (González *et al.*, 2007). Para representar el crecimiento de los animales, se han utilizado distintas funciones matemáticas que relacionan el cambio del peso en función de la edad, entre ellas se encuentran las funciones asintóticas de patrón sigmoideal: Brody, Richards, Logística, Gompertz y Von Bertalanffy (Fitzhugh, 1976); estas funciones matemáticas sigmoideas resumen el fenómeno del crecimiento en pocos parámetros, que pueden ser interpretados biológicamente: el parámetro "A", es una estimación del peso a la madurez, el parámetro " μ " está relacionado con la pendiente de la curva (la tasa de crecimiento o madurez) y el parámetro "B", es una constante de integración, existe un cuarto parámetro (T_0), el cual se relaciona con el punto de inflexión de la curva, (umbral entre la fase autoaceleradora y la fase autoinhibidora) (Vargas y Ulloa, 2008). En las regiones tropicales, la producción de novillas de reemplazo ha sido señalada como una de las etapas críticas y de mayor importancia en la producción de ganado, debido a las bajas ganancias de peso que experimentan los animales en el período posdestete, con el consiguiente retardo en los procesos de crecimiento y madurez fisiológica (Mukassa-Mugerwa, 1989). Entre los factores que modifican el inicio de la etapa reproductiva de la vaquilla son la edad y el peso; Además existen otros factores que influyen en el inicio de la etapa reproductiva, entre los que se encuentran el genotipo, el bajo estado nutricional de los animales, manejos inadecuados de salud, estacionalidad de los forrajes, factores de tipo social entre otros (Garverick y Smith, 1993). El objetivo del presente trabajo, fue determinar la curva de crecimiento de diferentes genotipos de novillas y ubicar en ella, la edad y el peso en la que se realiza la primera gestación, con la finalidad de obtener modelos que estimen el futuro comportamiento reproductivo de las becerras.

Material y Métodos:

La Posta Zootécnica "Torreón del Molino" de la Universidad Veracruzana (PZTM-FMVZ; UV), está localizada en el municipio de Veracruz, Ver., se ubica geográficamente en los 19° 10' de latitud Norte y 96° 10' longitud Oeste, a una elevación promedio de 15 msnm. El

grupo genético predominante es la cruce de *Bos Taurus* / *Bos Indicus* en diferentes proporciones de sangre. Los pastos con los que cuenta son: *Cynodon plectostachyus* (33%), *Digitaria decumbens* (30 %), *Brachiaria decumbens* (1.5%), *Brachiaria mítica* (9.4%), *Bachiaria brizantha* (5.1%) y *Cynodon dactylon* y *Paspalum dilatatum* (14%), el clima de la región es tropical subhúmedo, se clasifica como AW₁ (García, 1988). Los pesos vivos de las hembras se registran cada mes (desde el nacimiento). Alrededor de los 300kg las becerras son revisadas rutinariamente dos veces al día para la detección de los signos de estro (receptibilidad sexual) para así, ser inseminadas de forma artificial o por monta natural. Las razas que conforman los grupos genéticos del hato de doble propósito son: Holstein, Suizo, Cebú; de los cuales se integraron los siguientes grupos: Grupo 26-50% de sangre europea (n= 18), Grupo 51-75% de sangre europea (n=104), Grupo 76-100% de sangre europea (n= 10); N=132 animales. Las curvas de crecimiento se modelaron con Logistic y Gompertz (Nešetřilová, 2005; Altamirano *et al.*, 2006), con el uso del software Origin 7.0 y las graficas se realizaron con ayuda del software Sigma Plot 11.0.

Modelo Logistic:

$$P = \frac{a}{1 + e^{-\mu(T-B)}}$$

Donde:

P = Peso vivo (Kg)

A = Peso a la madurez (Kg)

T = Tiempo (Días)

B = Edad al punto de inflexión (Días)

μ = Velocidad de crecimiento (peso/ tiempo)

e = Base de logaritmo natural

Modelo Gompertz

$$P = ae \left(-e \left(\frac{T-T_0}{B} \right) \right)$$

Donde:

P = Peso vivo (Kg)

A = Peso a la madurez (Kg)

T = Tiempo (Días)

T_0 = Tiempo cero

B = Edad al punto de inflexión (Días)

e = Base de logaritmo natural

Se analizo la edad y el peso en el que las novillas se gestan por primera vez por ANDEVA del módulo GLM del paquete estadístico STATISTICA V7.0; se evaluó la discrepancia entre el peso observado y el peso esperado con el modelo matemático, el modelo estadístico fue el siguiente:

$$Y_{ij} = \mu + G_i + \varepsilon_{ij}$$

Donde:

Y_{ij} = Peso en cada genotipo

μ = media general

G_i = Genotipos $i = 1, 2, 3$; (26-50%, 51-75%, 75-100%)

ε_{ij} = Error experimental $\varepsilon_{ij} \sim NI(0, \sigma^2)$

Resultados y Discusión

Las curvas de crecimiento se han utilizado con diversos objetivos, uno de ellos es la definición de curvas estándares para las distintas razas o grupos genéticos, al respecto, Padrón y Vaccaro (1987), proponen que la cuantificación del crecimiento en peso hasta una edad determinada bajo condiciones óptimas de manejo, alimentación, control sanitario y condiciones climáticas en las cuales se ha desarrollado una raza bovina, constituye el patrón de crecimiento normal de la raza o grupo genético; también pueden utilizarse para determinar o predecir si un animal es excepcional o si es necesario adoptar medidas

correctivas en la explotación. Los resultados en la Figura 1, muestran el análisis de la edad y peso de las novillas de la PZTM-UV; la cual muestra los ajustes (líneas continuas) de los modelos Logistic y Gompertz que representan de manera eficiente ($R^2 = >0.80$) el comportamiento de los pesos en las diferentes edades de los grupos de novillas. El sistema de manejo, la calidad y la disponibilidad de forraje, así como la adaptación de los grupos genéticos en las condiciones ambientales que predominan en la región agroecológica, juegan un papel crucial en la forma y la posición de la curva de crecimiento en el eje tiempo, lo que modifica así en el inicio de su etapa reproductiva (Garverick y Smith, 1993), para el inicio de la etapa reproductiva, es más importante el tamaño del animal que la edad del mismo, con base en que el cerebro demanda “reconocer” cierto desarrollo corporal mínimo para dar inicio con los mecanismos reproductivos. (Hall, 1995; Foster y Nagatani, 1999).

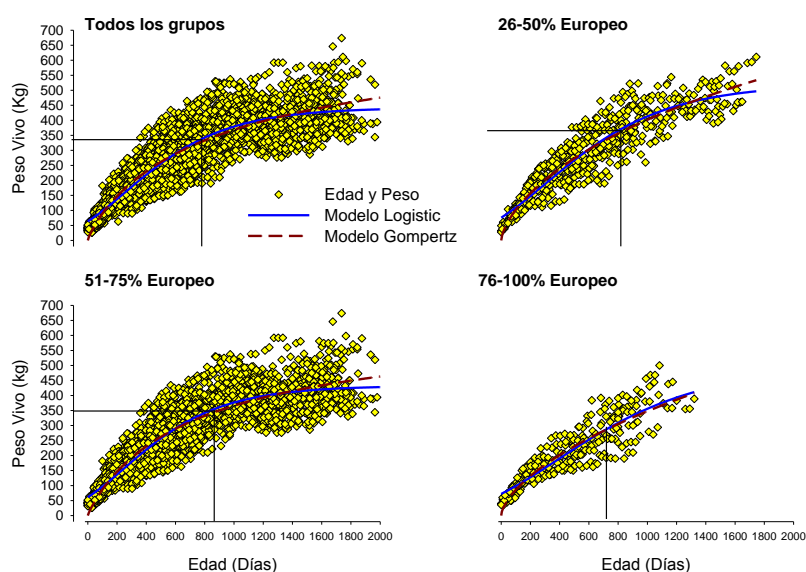


Figura 1. Curvas de crecimiento en las diferentes proporciones de sangre europea.

La línea continua (azul) representa el ajuste de los datos mediante el modelo Logistic, la línea continua punteada (Rojo) representa el ajuste de los datos mediante el modelo Gompertz; la línea que parte de los ajustes hacia el eje X, muestra la edad en la que los animales son gestados por vez primera, así como la línea que parte de los ajustes hacia el eje Y muestra el peso con el que los animales son gestados por vez primera.

En los parámetros de los ajustes, Altamirano *et al.* (2006) muestran valores muy similares a los obtenidos en el presente trabajo (Cuadro 1), con novillas en sistemas de doble propósito bajo condiciones de trópico húmedo. Altamirano *et al.*, (2006) reporta pesos a la madurez para el modelo Logistic que van de 455.3 kg hasta 511.7 kg (± 31.28 EEM) para grupos con diferentes proporciones de Holstein X Cebú que van de los 5/8H3/8C hasta grados de pureza como 59/64H5/64C. Vargas y Ulloa (2008) reportan un peso a la madurez con el modelo de Gompertz de 491.7 ± 27.10 Kg para cruza en las que predomina el Holstein sobre el Cebú, con un punto de inflexión de 276 días, parámetro menor al encontrado en el presente trabajo, y muy similar al reportado por Altamirano *et al.*, (2006)

que se encuentra en el orden de 415 días. Las novillas con proporción de sangre >75% de europeo tiene edades y pesos menores que en los otros grupos como lo han expresado otros autores (Galina y Arthur, 1989; Grajales *et al.*, 2006). Los coeficientes de correlación y de determinación comparados con los obtenidos por Altamirano *et al.*, 2006; Zarate y Vinay (2004) así como por Vargas y Ulloa (2008), son menores, muy probable a que se realizaron de manera grupal en los diferentes encastes de europeo de las diferentes proporciones genéticas dentro de cada grupo analizado. Bajo condiciones de trópico de Costa Rica, Hidalgo (1999) estimó que la edad al primer empadre está entre los 30 y 36 meses de edad (900-1080 días). La variación en la edad al inicio de su etapa reproductiva, indica que los animales responden acorde con las condiciones que se encuentran criados, manejo, fuentes y disponibilidad de alimento y circunstancias imperantes del medio; los valores del presente trabajo se muestran en el Cuadro 2. Adam y Robinson (1994) y McAndrews *et al.* (1993), consideran la hipótesis de que cuando un sujeto obtiene un "peso crítico", se dispara una serie de eventos endocrinos que inducen el inicio de su etapa reproductiva (pubertad) y el subsecuente desarrollo de una actividad sexual continua (Abeygunawardena y Dematawewa, 2004).

Cuadro 1. Parámetros obtenidos del ajuste de los datos mediante el modelo de Logistic en los grupos de sangre europeo.

| Modelos | Genotipos | Parámetros | | | | | |
|----------|-----------|----------------|----------------|------------------|----------------|------|----------------|
| | | A | B | μ | T_0 | R | R ² |
| Logistic | 26-50% | 491.07 ± 8.50 | 458.55 ± 14.83 | 0.0032 ± 0.00032 | | 0.92 | 0.84 |
| Gompertz | | 521.14 ± 11.6 | 472.69 ± 23.65 | | 309.88 ± 13.25 | 0.92 | 0.85 |
| Logistic | 51-75% | 418.01 ± 2.19 | 389.24 ± 4.51 | 0.00383 ± 0.0007 | | 0.91 | 0.83 |
| Gompertz | | 432.06 ± 2.74 | 376.99 ± 7.21 | | 252.54 ± 3.86 | 0.91 | 0.84 |
| Logistic | 76-100% | 438.33 ± 26.09 | 499.66 ± 48.15 | 0.0029 ± 0.00026 | | 0.89 | 0.80 |
| Gompertz | | 499.56 ± 45.71 | 575.15 ± 74.64 | | 378.82 ± 59.06 | 0.89 | 0.79 |
| Logistic | Todos | 425.80 ± 2.02 | 400.86 ± 4.38 | 0.0037 ± 0.00006 | | 0.91 | 0.83 |
| Gompertz | | 441.87 ± 2.79 | 393.93 ± 7.00 | | 260.96 ± 3.74 | 0.91 | 0.83 |

A= amplitud máxima (peso a la madurez), B= edad al punto de inflexión en días, μ = velocidad de crecimiento (peso/días), la cual muestra la pendiente de la curva, T_0 = tiempo cero, edad en la que da inicio la inflexión de la curva, R= coeficiente de correlación, R²= Coeficiente de determinación.

Cuadro 2. Valores estimados vs observados de la edad y peso a primera gestación.

| Grupos de Europeo | Modelo Logistic | | Modelo Gompertz | | Observados | |
|-------------------|-----------------|-----------------------------|-----------------|-----------------------------|-----------------------------|-----------------------------|
| | Edad (Días)* | Peso (Kg.) ^{&} | Edad (Días)* | Peso (Kg.) ^{&} | Edad (Días)* | Peso (Kg.) ^{&} |
| 26-50% | 813.42 | 372.55 | 817.96 | 370.43 | 819.04 ± 66.25 ^a | 381.44 ± 15.59 ^a |
| 51-75% | 781.35 | 341.85 | 777.30 | 336.96 | 785.22 ± 22.99 ^a | 347.44 ± 5.91 ^b |
| 76-100% | 738.86 | 293.87 | 727.58 | 289.57 | 727.14 ± 62.50 ^a | 336.50 ± 10.50 ^b |
| Todos | 796.05 | 345.12 | 785.15 | 339.24 | 788.00 ± 21.20 ^a | 351.5 ± 6.43 ^{ab} |

a y b, literales diferentes entre hileras de la misma columna son estadísticamente diferentes (p < 0.05) *, & Caracteres iguales entre columnas son estadísticamente iguales (p > 0.05)

Conclusión

Los Modelos Logistic y Gompertz, resultan eficientes para establecer predicciones del crecimiento y comportamiento productivo, reflejado en el futuro inicio de la actividad reproductiva de nuevas becerras candidatas a ser vientres de reemplazo.

BIBLIOGRAFÍA

- Abeygunawardena H. and Dematawewa C. M. B. (2004). Pre-pubertal and postpartum anestrus in tropical Zebu cattle. *Animal Reproduction Science* 82-83: 373-387.
- Adam C. L. and Robinson J. J. (1994). The role of nutrition and photoperiod in the timing of puberty. *Proceedings Nutrition Society* 53: 89-102.
- Altamirano B. R., Vinay V.J.C., Zárate M. J., López G.I., Villegas A.Y. (2006). Análisis del crecimiento de hembras Holstein y sus cruizas con cebú en el trópico. *Avances en la Investigación Agrícola, Pecuaria, Forestal y Acuícola en el Trópico Mexicano*. Libro científico número 3. Veracruz, Mexico. 435p.
- Fitzhugh H A. (1976). Analysis of growth curves and strategies for altering their shapes. *Journal of Animal Science* 42:1036
- Foster D.L., Nagatani S. (1999). Physiological perspectives on leptin as a regulator of reproduction: Role in timing puberty. *Biology of Reproduction* 60: 205-215.
- García E. (1988). Modificación al sistema de clasificación climática de Koppen (para adaptarlo a las condiciones de la republica mexicana) Offset Larios, S.A. 4ª edición. México, D.F. 217 pp.
- Garverick H.A., Smith M.F. (1993). Female reproductive physiology and endocrinology of cattle. *Veterinary Clinics of North America. Food Animal Practice*. 9: 223-247.
- Galina, C.S. y Arthur, G. H. (1989). Review of cattle reproduction in the tropics. Part 1. Puberty and age at first calving. *Animal Breeding Abstracts*. 57: 585-588.
- González V. D. M., Goicochea Ll., J., Quintero M. A., Rubio G.J.L., Aranguren M.J.A. (2007). Analisis de tres procedimientos estadísticos para la evaluación del crecimiento de mautas mestizas bajo diferentes regimenes nutricionales. *Revista Científica de la Facultad de Ciencias Veterinarias*. Universidad del Zulia, Facultad de Ciencias Veterinarias.
- Grajales H., Hernández A. Prieto E. (2006). Edad y peso a la pubertad y su relación con la eficiencia reproductiva de grupos raciales bovinos en el trópico colombiano. *Livestock Research for Rural Development* 18 (10) 2006.
- Hidalgo C. (1999) El futuro de la ganadería de carne en Costa Rica: La producción primaria. *Memoria XI Congreso Nacional Agronómico*. Costa Rica, 3: 539- 541
- Hall J.B., Staigmiller R.B., Bellows R.A., Short R.E., Moseley W.M., Bellows S.E. (1995). Body composition and metabolic profiles associated with puberty in beef heifers. *Journal Animal Science*; 73(11):3409-3420.
- McAndrews J. M., Stroud C. M., MacDonald R. D., Hymer W. C. and Deaver D. R.(1993). Age-related changes in the secretion of growth hormone in vivo and in vitro in infantile y prepubertal Holstein bull calves. *Journal of Endocrinology* 139: 307-315.
- Mukassa-Mugerwa E. (1989). A review of reproductive performance of female *Bos indicus* (zebu) cattle. *ILCA Monografía No. 6 International Livestock Centre of Africa* 117-125.
- Nešetřilová H. (2005). Multiphasic growth models for cattle. *Czech Journal Animal Science*, 50 (8): 347–354
- Padrón G. M. y Vaccaro R. (1987). Crecimiento de hembras pardo suizas bajo manejo intensivo. *Zootecnia Tropical* 5(1 y 2):77-93
- Vargas B y Ulloa J. (2008). Relación entre curvas de crecimiento y parámetros reproductivos en grupos raciales lecheros de distintas zonas agroecológicas de Costa Rica. *Livestock Research for Rural Development* 20 (7) s/p.
- Zarate M.J.P. y Vinay, V.J.C. (2004). Evaluación de dos modelos matemáticos para determinar el crecimiento de las hembras en un sistema de lechería tropical. *Avances en la Investigación Agrícola, Pecuaria, Forestal y Acuícola en el Trópico Mexicano*. Libro científico número 1. Veracruz, México. 359-366p.

Estudio del Bajo Peso en Escolares de 6-13 Años en 7 Escuelas de la Ciudad de Chilpancingo, Gro., México

Vivian Sistachs Vega¹, Lucio Díaz González²

¹ Universidad de la Habana, Cuba. vivian@matcom.uh.cu

² Universidad Autónoma de Guerrero, Méx. luciodiaz1@yahoo.com.mx

1. Introducción

Algunos estudios en México sostienen que el problema de desnutrición en el país recae más en el lado distributivo que en el de la producción (Lustig, 1984, p. 439). Esto no quiere decir que un incremento en la producción de alimentos no sea necesario en la lucha contra la desnutrición, de hecho es esencial, pero previene del supuesto erróneo de creer que elevar la producción de alimentos será suficiente para combatir el problema de la desnutrición.

De acuerdo con diversas estimaciones, se ha considerado que la desnutrición en el país es relativamente grave, destacando que el porcentaje global de desnutrición es de casi 30% de la población infantil menor de 5 años, de la cual casi la mitad se encuentra en situación de desnutrición grave; que la incidencia de bajo peso en los estados de la región sur del país (principalmente Guerrero, Oaxaca y Chiapas) es 2.2 veces mayor que en el norte y centro de la República Mexicana, y que el 77% de las personas con alto grado de desnutrición y pobreza se encuentran en el medio rural (dif-sep, 1993).

La magnitud de este problema y su incidencia a lo largo de todo el país ha conducido históricamente a la instrumentación de diversos programas para abatirlo, los cuales han sido muy diversos en magnitud y extensión. Sin embargo, a pesar de todos los esfuerzos realizados, los programas instrumentados hasta ahora muestran graves limitaciones de las instancias públicas en materia de su planeación, coordinación, capacidad de gerencia, operación, seguimiento y expansión de los mismos. Esta situación ha derivado en marcadas deficiencias, como por ejemplo la falta de delimitación de objetivos, y poblaciones a beneficiar, que ha derivado, a su vez, en esquemas poco focalizados y diferenciados respecto al valor nutricional del paquete alimentario, que debería estar dirigido a cada uno de los grupos atendidos.

En el año 2004 se llevo a cabo un estudio descriptivo de corte transversal para investigar el problema de la obesidad en los escolares en 7 escuelas de la Ciudad de Chilpancingo, Gro. México. En este estudio se encontró con el problema del bajo peso de los escolares que fue de aproximadamente 45%. El presente trabajo tiene como objetivo realizar un análisis exploratorio para conocer más acerca de este problema e investigar posibles factores asociados con este problema, para ello se aplicó análisis de correspondencia aplicado a la alimentación, actividad física y medidas antropométricas. Regresión logística para investigar la relación de estos factores con el bajo peso.

2. Material y Métodos

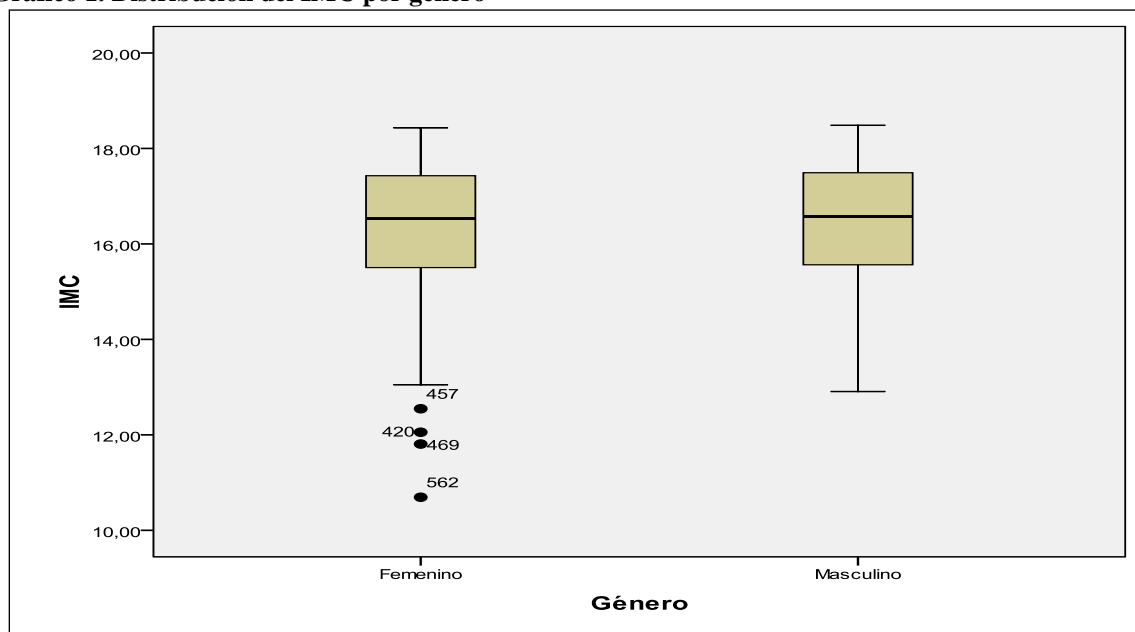
El estudio realizado fue observacional, descriptivo en 7 escuelas primarias públicas de Chilpancingo, Gro. México. Se diseñó y aplicó un cuestionario dividido en siete secciones: datos generales, datos socio demográficos, consumo de alimentos (alto riesgo, bajo riesgo y protectores), características y hábitos de los escolares, actividad física y antropometría.

Los criterios de inclusión utilizados para seleccionar a los participantes del estudio fueron: edad en el intervalo de 6-13 años, alumnos con bajo peso y normales. El estado nutricional se determinó en base al peso medido y peso esperado de acuerdo a su edad.

3. Resultados

En el estudio realizado se tiene un total de 643 escolares, de los cuales 326 (50.7%) son niñas y 317 son niños (49.3). El peso y la talla fueron de 33.26 y 132.14 respectivamente. Para este estudio se cuenta con 322 escolares con bajo peso y 321 con peso normal de acuerdo con su IMC, el 49% los escolares con bajo peso son mujeres y el 51% son hombres, no se encontró diferencias por género aunque se observan algunos valores atípicos que corresponden a niñas con un IMC bajo respecto a las demás ver gráfico 1. El IMC determinado por turno no presenta diferencias.

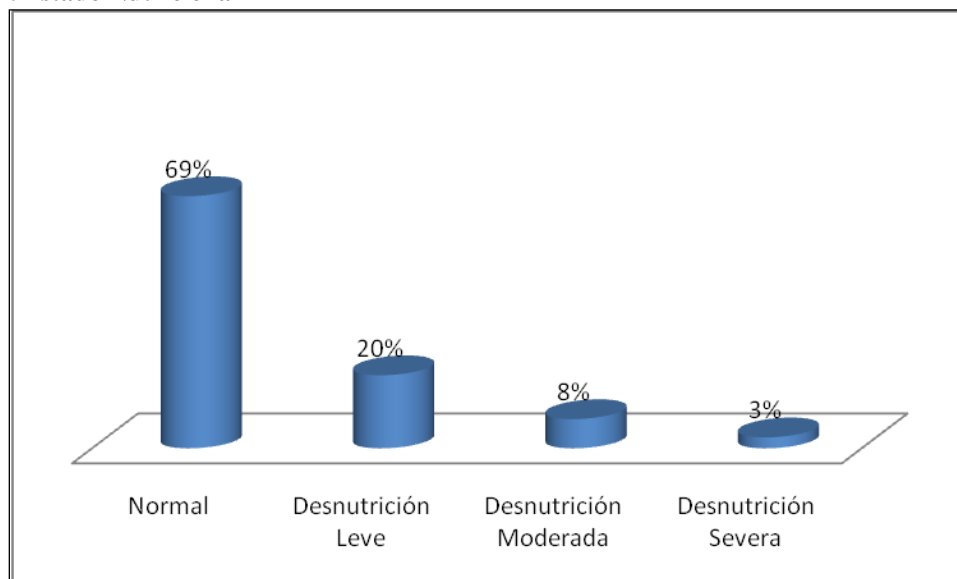
Gráfico 1. Distribución del IMC por género



Fuente: Encuesta realizada a los escolares en Chilpancingo, Gro. México, (2004)

Para determinar el estado nutricional de los escolares se dividió el peso real entre el peso esperado y se encontró que el 20% de los escolares con bajo peso tiene una desnutrición leve, el 8% moderada y 3% severa, ver gráfico 2.

Gráfico 2. Estado Nutricional



Fuente: Encuesta realizada a los escolares en Chilpancingo, Gro. México, (2004)

Análisis de regresión logística

Para determinar la prevalencia de los factores asociados al bajo peso se tomaron en cuenta algunas de las variables que tuvieron importancia en el estudio nutricional realizado como: edad, IMC (como variable dependiente con las categorías de normal y sobrepeso), peso al nacer, Ingreso económico de los padres, escolaridad del padre y de la madre, IMC del padre y de la madre, sexo, los resultados se muestran en las tablas 1 y 2.

Tabla 1. Tabla de clasificación

| Observado | | Pronosticado | | |
|-------------------|-------------|--------------|-----------|---------------------|
| | | Peso Normal | Sobrepeso | Porcentaje correcto |
| Estado | Peso Normal | 277 | 79 | 77,8 |
| | Sobrepeso | 83 | 217 | 72,3 |
| Porcentaje global | | | | 75,3 |

Fuente: Encuesta realizada a los escolares en Chilpancingo, Gro. México, (2004)

Tabla 2. Resultados Análisis de Regresión

| | B | S.E. | Wald | df | Sig. | Exp(B) |
|-------------------|--------|-------|---------|----|------|---------|
| EDAD | -,391 | ,053 | 53,826 | 1 | ,000 | ,676 |
| IMCMadre | -,116 | ,044 | 6,896 | 1 | ,009 | ,891 |
| IMCPadre | -,075 | ,044 | 2,888 | 1 | ,089 | ,928 |
| Género | -,317 | ,193 | 2,704 | 1 | ,100 | ,729 |
| Escolaridad madre | -,038 | ,031 | 1,548 | 1 | ,213 | ,963 |
| Escolaridad padre | ,059 | ,031 | 3,503 | 1 | ,061 | 1,061 |
| Ingreso Total | ,000 | ,000 | 6,527 | 1 | ,011 | 1,000 |
| Peso al Nacer | -,004 | ,000 | 78,393 | 1 | ,000 | ,996 |
| Constante | 19,865 | 1,923 | 106,760 | 1 | ,000 | 4,239E8 |

Fuente: Encuesta realizada a los escolares en Chilpancingo, Gro. México, (2004)

En esta tabla podemos observar que las variables que salieron significativas son el peso al nacer, la edad y el IMC de la madre, la interpretación que se le puede dar es que a menor peso al nacer, antecedente de madre con bajo peso y edad corta el niño es más propenso a tener un bajo peso.

4. Conclusión

La proporción de desnutrición de leve a severa es del 31% en niños escolares de la Ciudad de Chilpancingo. En el consumo de alimentos de alto riesgo en más del 80% de los niños tenemos: jamón, salchicha; de los alimentos de consumo diario el pollo y el chorizo son los de mayor consumo, también hay que destacar el consumo las papas fritas, chocolate en polvo, gelatina de agua, azúcar, así como el refresco.

En el consumo de alimentos protectores hay que destacar el arroz, pasta para sopa, maíz y frijol, hay que mencionar que la fruta se consume con mucha frecuencia en este grupo de niños con problemas de bajo peso.

No se encontró diferencia en el IMC por género ni turno. Al aplicar el análisis de regresión se encontró los escolares con madre con IMC bajo, bajo peso al nacer y edad corta son más propensos a tener un bajo peso.

REFERENCIAS

Instituto Nacional de Salud Pública. Encuesta Nacional de Nutrición 1999. Niños menores de 5 años. Tomo I. Cuernavaca, Mor. INSP, 2000.

Instituto Nacional de la Nutrición. Encuestas Nutricionales en México. México: División de Nutrición, 1963.

Iturbide Galindo, Laura; Rodríguez Arana, Ricardo; Olea, Edgar Gonzalo. La desnutrición infantil en México: Una propuesta de medición. *En publicación: Economía: Teoría y Práctica, NUEVA ÉPOCA, no. 9*. DCSH, Division de Ciencias Sociales y Humanidades, UAM, Universidad Autonoma Metropolitana: Mexico. 1998.

Levi S, Rivera DJ, Ramírez SC. Sobrepeso y obesidad en población escolar mexicana. Libro de resúmenes del IX Congreso Nacional de Investigación en Salud Pública, 2001: 34-35.

Instituto Nacional de Estadística, Geografía e Informática. Encuesta Nacional de Ingresos y Gastos de los Hogares, México: INEGI, 2000.

Hernández B, Gortmakers SL, Goldiestz GA. Association of obesity with physical activities, television programs and other forms of video viewing among children in Mexico City. *Int Obesity* 1999; 23: 845-54.

Rojas M, Guerrero L. Nutrición clínica y gastroenterología. Bogotá, Colombia, Médica Panamericana, 1998.

Sheldon WH, Stevens SS, Tucker WB. The varieties of human psyche an induction to constitutional psychology. EUA: Harper & Brothers, 1940.

Lusting, N. "La desigual distribución del Ingreso y la Riqueza en México, en R. Cordera y C. Tello (eds), La Desigualdad en México, Siglo XXI, 1984.

Inventario de Maguey Papalote (*Agave cupreata Trel et Berger*) de la Región Centro del Estado de Guerrero.

L. Alonso¹, D. Covarrubias², F. Maradiaga³

¹ Maestría en Estadística Aplicada. UAGro., Méx. alonso_lore@yahoo.com.mx

² Universidad Autónoma de Guerrero. Méx. dcova@uagro.mx

³ Centro de investigación área Ciencias Naturales. UAGro., Méx.

Resumen

El Maguey, es una planta monocotiledónea perteneciente a la familia agavaceas, misma que cuenta con mas de 273 especies. Actualmente existe un importante potencial productivo de Maguey en la región centro del Estado de Guerrero, la especie que abunda es la *cupreata Trel. et Berger* conocida regionalmente como Maguey Papalote. El Gobierno del Estado interesado en conocer el estado que guarda la producción del mezcal en la región para determinar las estrategias de apoyo a los productores, convino con la Universidad Autónoma de Guerrero en la realización de dicho inventario. Se presentan los resultados obtenidos en en la realización del inventario así como una estrategia de muestreo alternativa.

Palabras Clave: Inventario, muestreo por transectos .

1. Introducción

A nivel internacional, las especies de flora y fauna están severamente amenazados por estrategias de crecimiento económico desvinculadas de la conservación de los recursos naturales. Desde hace más de una década, se ha ido complementando con una visión ecológico-ambiental e integral del recurso natural, incluyendo además una valoración del recurso no maderable y de los servicios ambientales (agua,

biodiversidad, captura de carbono, ecoturismo), entre otros (FAO 1995).

Así, diferentes sectores gubernamentales y sociales en México y otros países, han tenido la necesidad de estudiar las relaciones entre los sistemas naturales y la sociedad, para detectar, medir, inventariar, monitorear y evaluar cambios en los ecosistemas en respuesta a las acciones humanas; lo que exige información y conocimiento espacio-temporal, creíble y defendible, esto para entender las relaciones causa-efecto de los cambios en los sistemas ecológicos y poder anticipar la solución a problemas y situaciones que amenazan al ecosistema, (Bahamondez *et al* 2005; Gruijter *et al* 2006; Covarrubias 2007).

Los responsables de gestión y planificación de los recursos naturales y culturales necesitan una evaluación rápida de la extensión espacial y magnitud de los daños o beneficios, como los desastres ecológicos tales como inundaciones, incendios, la contaminación o la producción agrícola de un área. Por este motivo, es necesario contar con un sistema integrado que proporcione la información sobre los cambios en el medio ambiente y las causas de estos.

En cada uno de los ejemplos expuestos se hace necesario cuantificar los daños a la diversidad ecológica motivo por el cual se hace necesario el uso de una metodología que contenga estrategias

eficientes para la recolección de la información y el análisis estadístico. Gruijter *et al* (2006), Mostacedo y Fredricksen (2000), Shiver y Borders (1996), coinciden en señalar que los métodos estadísticos aplicados a estudios en ecología, permiten mostrar si los cambios en una población ecológica obedecen a factores físicos o son producidos al azar, y por tanto son explicados mediante variables bióticas, variables abióticas o por condiciones de prueba en laboratorio.

Kempton(2002), Rubio (2000) y Covarrubias (2007), refieren que el uso de muestreo en las investigaciones ecológicas es frecuente, esto debido a la imposibilidad técnica o económica de censar poblaciones distribuidas en áreas extensas.

1.1. Planteamiento del problema

El visible aumento de la demanda de mezcal se debe a la tendencia de mantener las tradiciones y las costumbres, además el gran auge que ha tenido el tequila ha abierto el mercado nacional e internacional, para los diferentes mezcales de México; particularmente la demanda en Guerrero del maguey papalote (*Agave cupreata Threl et Berger*).

Los problemas que enfrentan los productores de mezcal en Guerrero, es el desconocimiento a corto y mediano plazo de disponibilidad de magueyes debido a la carencia de un sistema de inventario y monitoreo ágil y confiable de las dinámicas de las poblaciones naturales de dicha especie.

1.2. Objetivo

Proponer una estrategia de muestreo alternativa a la aplicada para desarrollar el inventario del *Agave cupreata Threl et Berger*, que pueda proporcionar la información sobre el estado en que se encuentran y que contribuya a orientar las actividades productivas y de conservación hacia el desarrollo rural sustentable en la región centro del Estado de Guerrero.

1.3. Selección del Área de Muestreo

Se propuso realizar el muestreo para la Unión de Productores de Mezcal Xochialehualatl fueron: 4 sitios en el municipio de Chilpancingo de los Bravo, 4 en Eduardo Neri, 1 en Mochitlán y 3 en Tixtla de Guerrero, seleccionados bajo el criterio de la existencia de poblaciones silvestres de maguey

papalote, que fuesen áreas de aprovechamiento para la producción de mezcal.

Para la estimación de la densidad y estructura etaria de las poblaciones de maguey papalote se levantó información en 289 unidades primarias de muestreo circulares de 300 m² cada una, subdivididas en cuatro cuadrantes. Las unidades fueron distribuidas en los 12 sitios ver cuadro (1.3). Dichos sitios fueron elegidos a partir de reconocer las áreas de aprovechamiento de maguey papalote en la región centro.

Se empleó el método de muestreo en conglomerados sistemáticos monoetápicos ver Prodan *et al* (1997).

En todos los sitios de muestreo las unidades se ubicaron equidistantes (a 100 m) a lo largo de uno o más transectos. Además, se registraron las coordenadas UTM del centro de cada unidad de muestreo, con una precisión superior a los 10 m ver Maradiaga *et al* (2004). Lo cual es importante para georeferencia y modelaje espacial, así como para su futura localización e incluso, la ampliación del número de unidades de muestreo, de acuerdo al tiempo y financiamiento disponible.

Para establecer la estructura etaria y hacer la estimación de la capacidad de extracción, los especímenes fueron asignados dentro de siete categorías o etapas de desarrollo: Plántula, Juvenil menor, Juvenil medio, Juvenil mayor, Velilla, Capón y Calehual. Las últimas tres categorías corresponden a etapas donde la planta ha entrado en un proceso de maduración que, es cuando puede ser preparada para aprovecharse. Los resultados del inventario se presentan en el cuadro (1.3)

2. Estrategia alternativa

Un método alternativo es el método de los transectos que es ampliamente utilizado por la rapidez con se mide y por la mayor heterogeneidad con que se muestrea la vegetación. Un transecto es un rectángulo situado en un lugar para medir ciertos parámetros de un determinado tipo de vegetación. El tamaño de los transectos puede ser variable y depende del grupo de plantas a medirse.

| Localidad | UPM |
|------------------|-----|
| Axaxacualco | 53 |
| El Aguacate | 10 |
| Mazatlán | 10 |
| Mirabal | 12 |
| Mochitlán | 58 |
| Ojitos de Agua | 41 |
| Palo Blanco | 25 |
| Plan de Guerrero | 13 |
| Salto Valadez | 10 |
| Tlanipatla | 24 |
| Xochipala | 21 |
| Xocomanatlán | 12 |

Cuadro 1: Sitios muestreados

| Localidad | Promd Unidad | Prom hta |
|------------------|--------------|----------|
| Axaxacualco | 24.6 | 818.87 |
| El Aguacate | 49.9 | 1663.3 |
| Mazatlán | 25.1 | 836.7 |
| Mirabal | 7.0 | 233.33 |
| Mochitlán | 27.3 | 908.6 |
| Ojitos de Agua | 27.9 | 930.1 |
| Palo Blanco | 47.4 | 1578.7 |
| Plan de Guerrero | 23.5 | 782.1 |
| Salto Valadez | 39.0 | 1300.0 |
| Tlanipatla | 26.4 | 879.2 |
| Xochipala | 6.1 | 201.6 |
| Xocomanatlán | 10.5 | 350.0 |

Cuadro 2: Resultados

El transecto es muy utilizado en estudio de poblaciones bióticas (Vegetales y animales) para estimar la abundancia o la densidad (total de números divididos por el área) de individuos, en instancia se especifica planta o animal. Se mueven a lo largo de una o más líneas seleccionadas aleatoriamente, y la localización de distancias perpendiculares de cada individuo de la especie detectada.

Línea de transectos para estimar el total

$$\hat{t}_{ns}(z) \sum_{i=1}^n z_i \frac{B}{n2w} = \frac{B}{2w} \bar{z}_s \quad (1)$$

La densidad es estimada por

$$\bar{z}_{ns} = \frac{1}{A} \hat{t}_{ns}(z) \quad (2)$$

un estimador preferible en caso de diferencias largas en la longitud de la franja, el rango estimado es:

$$\bar{z}_R = \frac{\frac{B}{2w} \sum_{i=1}^n z_i}{\sum_{i=0}^n L_i} \quad (3)$$

La varianza de puede ser estimada por

$$\hat{V} \{ \hat{t}_{ns}(z) \} = \left(\frac{B}{2w} \right)^2 \frac{S^2(\widehat{Z}_i)}{n} \quad (4)$$

donde $S^2(\widehat{Z}_i) = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z}_s)^2$ La varianza de la densidad puede ser estimada dividiendo la varianza estimada del total estimado por A^2 .

El transecto también puede ser seleccionado longitudinal con muestreo probabilístico proporcional (pps). Este puede ser implementado seleccionando n localidades aleatoriamente de un área objetivo y se calculan líneas a través de puntos perpendiculares de la línea-base; esto es:

$$\hat{z}_{pps} = \frac{1}{n} \sum_{i=1}^n \hat{z}_i = \frac{1}{n} \sum_{i=1}^n \frac{z_i}{2wL_i} \quad (5)$$

y su varianza

$$\hat{V}(\hat{z}_{pps}) = \frac{1}{n(n-1)} \sum_{i=1}^n (z_i - \hat{z}_{pps}) \quad (6)$$

En los transectos, generalmente se miden parámetros como altura de la planta, abundancia, densidad y frecuencia.

La abundancia de Maguey silvestre en el Estado de Guerrero, así como el alto grado de participación social, donde los términos técnicos, agrícolas y agroindustriales, así como económicos apenas empiezan a introducirse y a ser utilizado por los técnicos y las organizaciones de productores más desarrolladas.

Referencias

- [1] Aguirre D. X. (2004). "Genética de poblaciones de *Agave cupreata* y *Agave potatorum*: aportaciones para el manejo y la conservación de dos especies mezcaleras". Lab. de Evolución Molecular y Experimental Instituto de Ecología, UNAM, México.
- [2] Bahamondez C., Lorenz M., Mery G. y Varjo J. (2005). "Forest Assessment for Changing Information Needs". IUFRO World Series, España.

- [3] Gentry H. (1982). "Agaves of continental North America". The University of Arizona Press, Tucson, Arizona, U.S.A.
- [4] Gruijter J., Brus D., Bierkens M. and Knotter M. (2006). "Sampling for Natural Resource Monitoring". Springer, Berlin.
- [5] Maradiaga C. F. S. (2004). *Desarrollo de un Sistema de Inventario y Monitoreo de Maguey Papalote (Agave cupreata Trev. & Berger) en el estado de Guerrero*. Fundación PRODUCE Guerrero A.C., Programa de Recursos Biológicos Colectivos (CONABIO) e Instituto de Investigación Científica Área Ciencias Naturales de la UAGro. Chilpancingo, Gro. México.
- [6] Mostacedo B., Fredericksen T. S., BOLFOR (2000). "Manual de Métodos Básicos de Muestreo y Análisis en Ecología Vegetal". El País, Santa Cruz, Bolivia.
- [7] Prodan, M., R. Peters, F. Cox y, P. Real. (1997). *Mensura Forestal IICA - GTZ*. San José, Costa Rica.
- [8] Ramírez A. (2006). "ECOLOGÍA. Métodos de muestreo y análisis de poblaciones y comunidades". Pontificia Universidad Javeriana Colección Biblioteca del Profesional
- [9] Rubio D. A. (2000). "Evaluación del Tamaño de la Parcela en Inventarios Forestales por Muestreo". Departamento de Estadística e Informática UNALM, México.
- [10] Shiver D. B. and Borders B. E. (1996). "Sampling Techniques for Forest Resource Inventory". John Wiley & Sons, Inc. New York.
- [11] FAO (1995). "Food and Agriculture Organization of the United Nations". Non-wood Forest Products for Rural Income and Sustainable Forestry. Non-wood Forest Products 7. FAO, Rome.

Estudio Comparativo entre Algoritmos de Clasificación Automática usando Simulaciones con el Paquete WEKA

Capilla, J.¹, Reyes, H.², Linares, G.³

¹ Facultad de Ciencias Físico Matemáticas. BUAP. hreyes@fcfm.buap.mx

² Facultad de Ciencias de la Computación. BUAP. jse_santiago@hotmail.com

³ DICA. BUAP. gladys.linares@icbuap.buap.mx

Resumen

Se presenta un estudio de comparación entre los algoritmos de agrupamiento EM y K-medias, que se encuentran implementados en el software libre WEKA acrónimo de Waikato Environment for Knowledge Analysis. El propósito es encontrar diferencias significativas entre estos métodos de agrupamiento. Los datos se simulan con el software estadístico R, en donde se varía el tipo de distribución, el tamaño de muestra, el número de variables y el número de cluster. Tomando como covariables número de variables, tamaño y número de cluster. En esta etapa preliminar se concluye que el algoritmo con mejor desempeño es el EM.

I. Introducción

En la actualidad con los progresos computacionales se han creado bases de datos de gran volumen. Unos ejemplos de estos casos son: La base de datos de Walmart añade aproximadamente un billón de registros diariamente (USA TODAY., 2006); Digital Palomar Sky Survey (DPOSS) genera imágenes de aproximadamente 3TeraBytes de píxeles por día (AAS., 2002) o también la red Yahoo que transfiere cerca de 10 Terabytes de datos por día, que es utilizada por alrededor de 410 millones de usuarios en un mes, lo cual también genera una gran explosión de información (Piatestsky G., 2006). Por esta razón es importante estudiar métodos eficientes que permitan encontrar patrones ocultos en enormes bases de datos.

II. Algoritmos de Agrupamiento (clustering)

En este procedimiento, la tarea es dividir los datos en grupos de objetos parecidos, que utilizan diferentes métricas de distancia para poder medir la similitud entre ellas, entre las cuales tenemos: la distancia Euclidiana, la Manhattan, la Mahalanobis, entre otras. El clustering es principalmente una técnica de aprendizaje automático y juega un papel importante en minería de datos, como es la exploración de datos científicos, recuperación de información, aplicaciones Web, marketing, análisis de ADN en biología computacional, etc. Se presenta a continuación, los dos métodos para el estudio, desde el punto de vista de WEKA.

(a) El algoritmo EM en WEKA

El algoritmo asigna una distribución de probabilidad a cada caso, lo que indica que la probabilidad de que pertenezca a cada una de las agrupaciones. EM puede decidir el número de grupos para la validación cruzada o se puede especificar cuántos grupos se quiere generar.

La validación cruzada (CV), determina el número de grupos mediante los siguientes pasos:

1. El número de grupos se establece en 1
2. El conjunto de prueba se divide al azar en 10 pliegues.

3. Algoritmo EM se realiza 10 veces utilizando los 10 pliegues de forma habitual de CV.
4. Loglikelihood es la media de los 10 resultados.
5. Si Loglikelihood aumenta el número de grupos se incrementará en 1 y el programa continúa en el paso 2.

El número de pliegues se fija en 10, mientras el número de casos en el conjunto de formación no es menor de 10. Y sus opciones de configuración son:

- Debug - Si es asignado true, puede salir información adicional a la consola del cluster.
- DisplayModelInOldFormat – Da la posibilidad de elegir entre el formato nuevo y el viejo. El formato viejo es mejor cuando hay muchos grupos. El nuevo formato es mejor cuando hay un menor número de grupos y muchos atributos.
- MaxIterations - número máximo de iteraciones
- MinStdDev - conjunto mínimo permitido de la desviación estándar
- NumClusters - número de agrupaciones. Si se deja -1 optimiza el número de grupos para la validación cruzada.
- Seed - El número de la semilla que se utiliza.

(b) Algoritmo K-Medias en WEKA.

Se trata de un algoritmo clasificado como Método de Particionado y Recolocación. El método de k-medias (Hartigan, J., 1975; Hartigan, J. & Wong, M., 1979), es hasta ahora el más utilizado en aplicaciones científicas e industriales (Garre et al., 2007).

La diferencia sobre este algoritmo, se debe especificar el número de agrupaciones que se quiere obtener. Y ese es un problema ya que se debe elegir un número óptimo.

Las opciones en WEKA son:

- displayStdDevs - Muestra las desviaciones std. de atributos numéricos y cuenta los atributos nominales.
- dontReplaceMissingValues - Sustituir los valores a nivel global con la media o la moda.
- numClusters - número de agrupaciones.
- seed - semilla de número aleatorio que se utiliza, (WEKA., 2008).

III. Metodología

Los datos que se usan para la comparación se crean en forma aleatoria usando el programa R, para dos tipos de distribuciones:

- Normal Estándar (0,1).
- Uniforme.

Además, el número de variables se toma de 3 a 7, en incremento de 1 y cada variable tiene 200 datos. Después se exportan los datos para la ejecución en WEKA, el cual corre con archivo de extensión csv con la cantidad de datos ya mencionados

IV. Resultados

El diseño resultante es: 2(distribuciones)*2(Métodos)*5(variables, de 3 a 7 en incremento de 1) y covariables (Nºclus y Tamaño), con un total de 10,000 datos creados para la simulación.

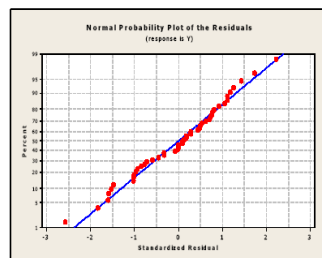
La tabla de ANOVA(al ejecutarla sobre Minitab 15), da los siguientes resultados: la distribución, número de variables y el método no afectan los resultados de la varianza y ya que están por debajo del $\alpha=,5$.

Se ve claramente que el método y la distribución (normal y uniforme) tienen una relación muy baja en la agrupación. Por tanto es no significativa.

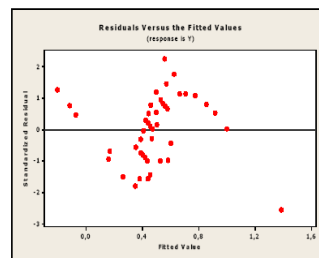
| General Linear Model: Y versus Distribución; Método | | | | | | |
|---|-----------|---------------|---------|---------------------|-------|-------|
| Factor | Type | Levels | Values | | | |
| Distribución | fixed | 2 | 1; 2 | | | |
| Método | fixed | 2 | 1; 2 | | | |
| Analysis of Variance for Y, using Adjusted SS for Tests | | | | | | |
| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
| Nvar | 1 | 0,00196 | 0,02035 | 0,02035 | 0,63 | 0,434 |
| Tamaño | 1 | 1,16938 | 3,11593 | 3,11593 | 95,70 | 0,000 |
| Noclus | 1 | 1,92779 | 1,46329 | 1,46329 | 44,94 | 0,000 |
| Distribución | 1 | 0,02348 | 0,02348 | 0,02348 | 0,72 | 0,401 |
| Método | 1 | 0,01735 | 0,01735 | 0,01735 | 0,53 | 0,470 |
| Error | 42 | 1,36755 | 1,36755 | 0,03256 | | |
| Total | 47 | 4,50750 | | | | |
| S = 0,180446 | | R-Sq = 69,66% | | R-Sq (adj) = 66,05% | | |
| Term | Coef | SE Coef | T | P | | |
| Constant | 3,1820 | 0,3256 | 9,77 | 0,000 | | |
| Nvar | -0,01580 | 0,01998 | -0,79 | 0,434 | | |
| Tamaño | -0,017821 | 0,001822 | -9,78 | 0,000 | | |
| Noclus | -0,43791 | 0,06532 | -6,70 | 0,000 | | |

Tabla 1. ANOVA del estudio de simulación.

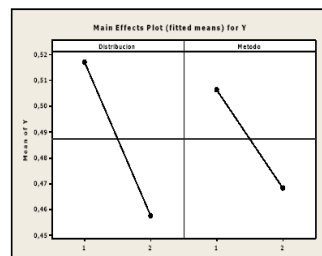
Se puede apreciar con mas facilidad en las Grafica 1 y 2, como se logra una estandarización en los residuos (normalidad) y homogeneidad de los datos resultantes, con respecto a la variable respuesta (Y), la cual contiene una distribución Binomial propuesta por (Von A., & Mair P., 2008). Pero hace falta independencia en los residuos. Y ya que estos son resultados preliminares, no afecta al modelo resultante al propósito de evaluar el funcionamiento de los dos métodos con las dos distribuciones.



Grafica 1.



Grafica 2.



Grafica3.

V. Conclusiones

Se llega a la consideración en el estudio que no importa el tipo de distribución (normal y uniforme) de los datos, ya que de igual manera se agrupan, pero si se tiene una mejora (no tan significativa) en el método EM sobre el método K-medias,

VI. Referencia

- AAS., (2002). Bulletin American Astronomical Society. Electronic Edition ISSN:0002-7537 20th Meeting Program, Albuquerque NM. Volumen 34, No.2, Sección 60.06.
<http://www.aas.org/publications/baas/v34n2/aas200/153.htm>. Activa (abril de 2006).
- Garre M., Cuadrado J., Sicilia M., Rodríguez D., Rejas R., (2007). Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software, Revista Española de Innovación, Calidad e Ingeniería del Software, Vol.3, No. 1
- Hartigan, J., (1975). Clustering Algorithms, John Wiley & Sons, New York.
- Hartigan, J. & Wong, M., (1979). "Algorithm AS139: A k-means clustering algorithm", Applied Statistics, Vol. 28, 100-108.
- Minitab Release 15, 2005. Statistical Software, Minitab. Inc.
- Piatestsky G., (2006). KDnuggets: "Data Mining, Knowledge Discovery, Text Mining Web Mining", Springer Science + Business Media B.V. ISSN: 1573-756X Online <http://www.kdnuggets.com/news/2005/n20/4i.html>. Activa en abril.
- R version 2.8.1, (2008). The R foundation for Statistical Computing, <http://www.r-project.org/>
- USA TODAY., (2006). USATODAY.com. a division of Gannett Co Inc. 7950 Jones Branch Drive, McLean VA22107.
http://www.usatoday.com/money/industries/technology/maney/2006-05-16-nsaprivacy_x.htm. Activa (abril de 2006).
- Vásquez B., (2009). Estudio comparativo de Métodos de clasificación Automática en la Zonificación Agro Ecológica del sur del estado de Puebla, tesis Maestría BUAP., Ciudad de Puebla.
- Von A., & Mair P. (2008). Evaluating cluster solutions with reference to data generation processes: A simulation study, Memorias del XXII Foro Nacional de Estadística, pp. 123-131. Aguascalientes, México: Instituto Nacional de Estadística, Geografía e Informática
- WEKA versión 3.5.8 (1999-2008). Desarrollado por la universidad de Waikato, Nueva Zelanda <http://www.cs.waikato.ac.nz/~ml/weka/>

Evaluación del Efecto del Tratamiento Térmico sobre los Carotenoides en Galletas Adicionadas con Espinaca Mediante Superficie de Respuesta

A. Reyes-Téllez¹, E.N. Aquino Bolaños², R.I. Guzmán³, I. Verdalet Guzmán⁴, D. Leyva Daniel⁵, M. Vázquez Hernández⁶, Z.J. Dominguez Esquivel⁷, E.R. Silva Hernández⁸

¹ Instituto de Ciencias Básicas. UV. arabellez02@gmail.com

² Instituto de Ciencias Básicas. UV. noraquino@yahoo.com.mx

³ Instituto de Ciencias Básicas. UV

⁴ Instituto de Ciencias Básicas. UV

⁵ Instituto de Ciencias Básicas. UV

⁶ Unidad S.A.R.A. UV

⁷ Unidad S.A.R.A. UV

⁸ Instituto de Ciencias Básicas. UV. esilva.basicas@gmail.com

Resumen

En los últimos años se afianza cada vez más el concepto de que los alimentos desempeñan un papel clave en la promoción de la salud, el bienestar y la reducción del riesgo de padecer determinadas enfermedades, lo que contribuyó a la aparición del concepto de “alimentos funcionales”. Este concepto está dirigido hacia nuevos avances en la nutrición que respalden y validen el desarrollo de nuevos alimentos y componentes alimentarios que proporcionen un beneficio adicional a la salud que van más allá de los efectos nutricionales básicos. Sin embargo; hay alimentos funcionales que ya existen en la naturaleza, un ejemplo de éstos es la espinaca, que contiene luteína, a la que se atribuye la reducción del riesgo de contraer diversos tipos de cáncer, enfermedades cardíacas y oftalmológicas además de ser utilizada como antioxidante en la industria alimentaria. La espinaca aporta el β -caroteno, mucho más conocido debido a su capacidad de transformarse en vitamina A en nuestro organismo. Una gran variedad de estudios indican que ambos compuestos actúan como antioxidantes y que son benéficos para la prevención de diversas enfermedades crónicas humanas no transmisibles. Una forma atractiva de diversificar el consumo de carotenoides puede ser a través de su incorporación en una galleta. Sin embargo, no se tienen referencias del efecto de las condiciones del horneado sobre el contenido de los carotenoides. Por lo que el objetivo de este trabajo fue determinar el efecto del tiempo (t) y temperatura (T) del horneado así como la cantidad de espinaca (e) adicionada sobre el contenido de luteína y β -caroteno. Para cumplir con este objetivo se utilizó un análisis de superficie de respuesta teniendo como factores la temperatura de horneado (250, 275 y 300 °F), tiempo (25, 30 y 35 min) y cantidad de espinaca (3,4 y 5 g) y como variables respuesta el contenido de luteína y β -caroteno. Se utilizó un modelo factorial fraccionado de 9 corridas y los resultados se analizaron utilizando el paquete estadístico Statistica 8.0. La ecuación que describe el comportamiento de la luteína en función de los tres factores analizados es $y_e = \beta_0 + \beta_1 e^2 + \beta_2 t^2 + \beta_3 T$ obteniendo un coeficiente de determinación $R^2 = 0.799$. El punto crítico de inflexión de los datos para el caso de la luteína/100 g de galleta fue de 3.97 g de espinaca, 20.67 min y 296 °F. La ecuación obtenida para el β -Caroteno/100 g galleta fue: $y_c = \beta_0 + \beta_1 e + \beta_2 t + \beta_3 T + \beta_4 T^2$. Presentando un coeficiente de determinación de $R^2 = 0.8617$. El punto crítico (mínimo) de los datos para el caso de β -caroteno/100 g galleta

corresponde a 4.77 g de espinaca, 32.91 min y 290.46 °F. El análisis de correlación de Spearman reveló que existen asociaciones significativas ($p \leq 0.05$) negativas entre el contenido de Luteína y la temperatura de horneado. Las mismas asociaciones anteriores se observaron con el contenido de β -caroteno. Las gráficas de superficie de respuesta sugieren una combinación de cantidad de espinaca de 4 g, tiempo de horneado de 35 min. y temperatura de horneado de 250°F como las condiciones óptimas para maximizar el contenido de luteína.

Introducción

Los carotenoides son un grupo de pigmentos que producen colores que van del amarillo al naranja y se dividen en a) carotenos (como el β -caroteno) y b) xantofilas (como la luteína). Una gran variedad de estudios indican que ambos compuestos actúan como antioxidantes y son benéficos para la prevención de diversas enfermedades crónicas humanas no transmisibles. Los carotenoides, además de sus propiedades nutricionales y de ser benéficos para la salud, son responsables del color atractivo de algunas frutas y hortalizas como zanahorias, brócoli, lechuga, col verde, judías verdes, perejil, calabaza, naranja, papaya y mango (Rodríguez-Amaya, 1997). Estos alimentos son relativamente fáciles de producir y disponibles prácticamente durante todo el año, son de bajo costo y por lo tanto fuentes accesibles de provitamina A para la mayoría de los países en desarrollo.

La espinaca (*Spinacia oleracea L.*) es una hortaliza de importancia nutrimental al ser una buena fuente de carotenoides y luteína. Una forma de diversificar el consumo de espinaca es a través de galletas. Sin, embargo, no se tienen referencias del efecto de las condiciones del horneado sobre el contenido de luteína y β -caroteno. De esta manera, el objetivo de este trabajo fue optimizar las condiciones de horneado (cantidad de espinaca, tiempo y temperatura) de galletas enriquecidas con espinaca para maximizar sus contenidos de luteína y β -caroteno.

Material y métodos

Para la elaboración de las galletas se utilizó harina integral (25 g), sal (0.125 g), polvo para hornear (0.010 g), margarina (7.5 g), agua (5 mL) y espinaca (3, 4 ó 5 g). Se mezclaron los ingredientes, se amasaron y la masa se laminó para formar las galletas. Para evaluar las condiciones de horneado se utilizó un análisis de superficie de respuesta teniendo como factores a la temperatura de horneado (250, 275 y 300 °F), el tiempo de horneado (25, 30 y 35 min) y la cantidad de espinaca (3, 4 y 5 g). Las variables respuesta fueron los contenidos de luteína y β -caroteno. Se utilizó un modelo factorial fraccionado de 9 corridas y los resultados se analizaron utilizando el paquete estadístico Statistica 8.0. La identificación y cuantificación de luteína y β -caroteno en las galletas se realizó mediante el método reportado por Panfili, et al. (2004).

Resultados

El análisis de superficie de respuesta para la maximización del contenido de luteína en las galletas elaboradas se muestra en la Figura 1. La parte roja de cada gráfica indica el valor máximo obtenido de luteína. El valor máximo de luteína se da, de acuerdo a los niveles empleados, con 4 g espinaca añadida y 35 minutos de horneado a 250 °F. El resultado sugiere que las temperaturas bajas de horneado no degradan la luteína y que al parecer el tiempo más prolongado favorece la salida de la misma de la célula vegetal. Es interesante el destacar que el contenido mayor de espinaca (5 g) no fue el que maximizaba el contenido de luteína, posiblemente debido a una mayor exposición al calor en la superficie de la galleta. La ecuación que describe el comportamiento de la luteína en función de los tres factores analizados es $y_e = \beta_0 + \beta_1 e^2 + \beta_2 t^2 + \beta_3 T$ obteniendo un coeficiente de determinación $R^2 = 0.799$, el cual indica un ajuste razonable de los datos al modelo sugerido.

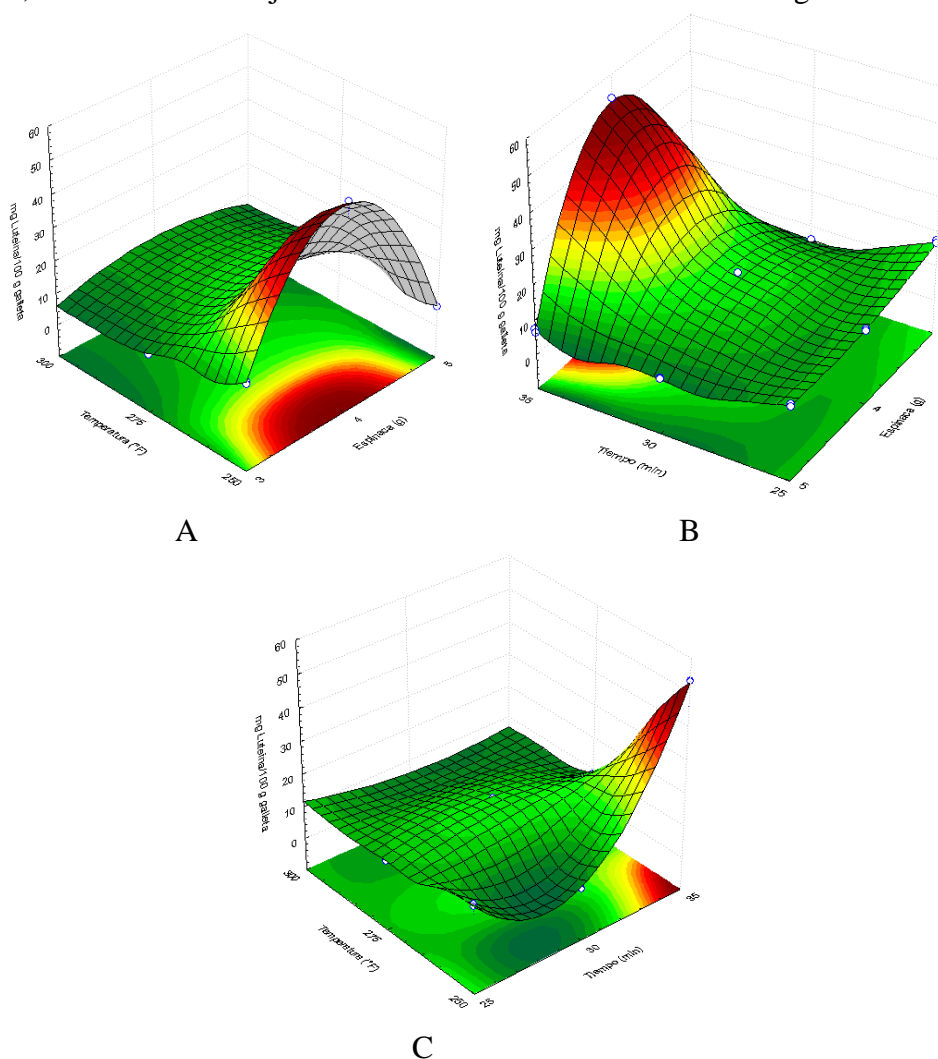


Figura 1. Gráficos de superficie de respuesta para la maximización del contenido de luteína (L) en galletas enriquecidas con espinaca de acuerdo a los parámetros “Contenido de espinaca” (E), “Temperatura” (T) y “Tiempo” (t). A: ExTxL. B: ExtxL. C: txTxL.

En lo referente al análisis de los carotenos (Figura 2), la temperatura baja y el tiempo corto favorecen la maximización del contenido de carotenos. Pareciera que existe un efecto protector del resto de los ingredientes de la galleta sobre los carotenos, ya que a menor cantidad de espinaca añadida a la galleta, mayor cantidad de carotenos se encontraba en la misma. El valor máximo de carotenos se alcanza con 3 g espinaca añadida y 25 minutos de horneado a 250 °F. La ecuación obtenida fue: $y_c = \beta_0 + \beta_1 e + \beta_2 t + \beta_3 T + \beta_3 T^2$ con un coeficiente de determinación de $R^2 = 0.8617$. Este último valor puede considerarse razonablemente bueno como para concluir que los datos sí se ajustan al modelo planteado.

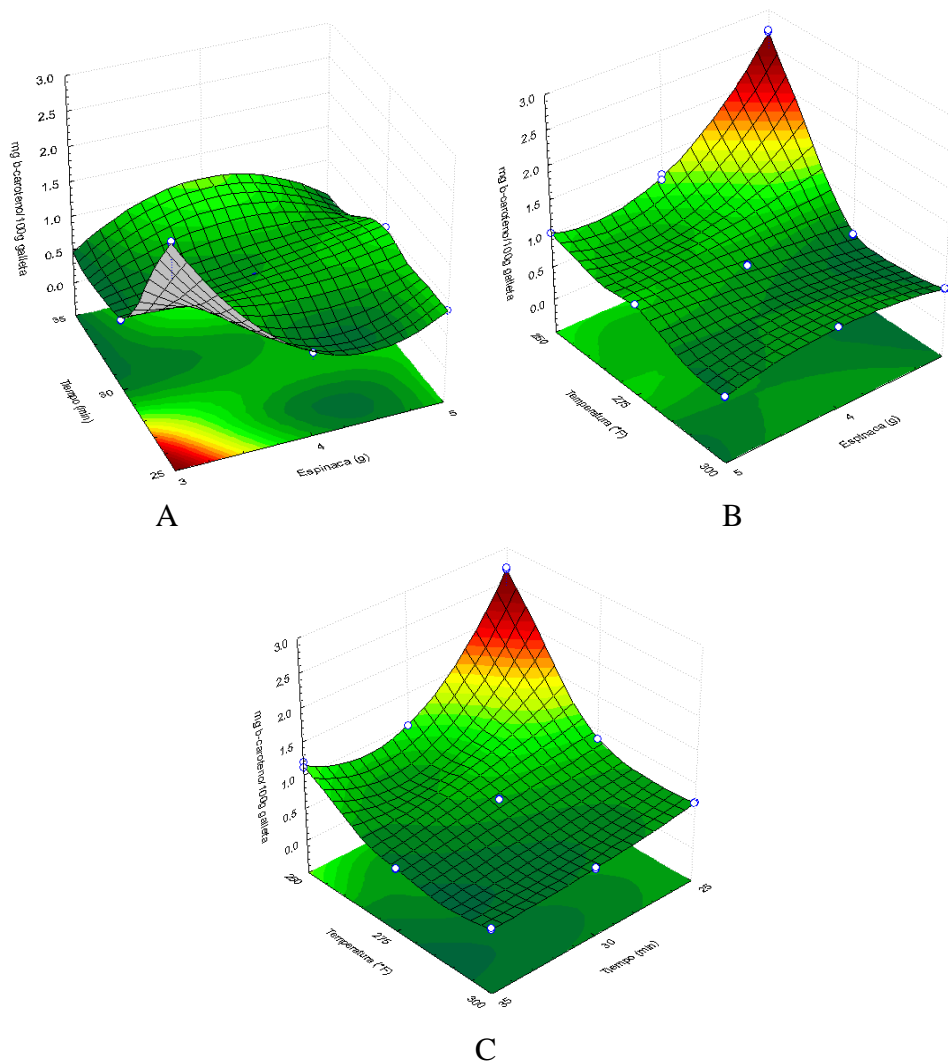


Figura 2. Gráficos de superficie de respuesta para la maximización del contenido de carotenos (Cr) en galletas enriquecidas con espinaca de acuerdo a los parámetros “Contenido de espinaca” (E), “Temperatura” (T) y “Tiempo” (t). A: ExTxCr. B: ExtxCr. C: txTxCr.

Conclusión

Debido a que el contenido de luteína es considerado de mayor importancia que el de los carotenos, se decidió considerar una combinación de cantidad de espinaca de 4 g, tiempo de horneado de 35 min. y temperatura de horneado de 250°F como las condiciones óptimas para maximizar el contenido de luteína en las galletas en estudio. La metodología de Superficie de Respuesta fue fundamental para la optimización de los contenidos de luteína y carotenos en galletas enriquecidas con espinaca, ya que desde el diseño hasta el análisis del mismo, se empleó menor cantidad de ingredientes y se ahorró tiempo valioso.

REFERENCIAS

Panfili, G., Fratianni, A., and Distaam, M.I. (2004). Improved Normal-Phase High-Performance Liquid Chromatography Procedure for the Determination of Carotenoids in Cereals. *Journal of Agriculture and Food Chemistry* **52**, 6373-6377.

Rodriguez-Amaya, D. (1997). *Carotenoides y preparación de Alimentos: La retención de los carotenoides provitamina A en los alimentos preparados, procesados y almacenados*. Campinas, Brasil: Departamento de ciencias de los alimentos. Universidad Estadual de Campinas, Brasil.

Estudio de la Liberación de Sabor en Goma de Mascar, Utilizando Microcápsulas de Aceite Esencial de Menta

S. C. Pereyra¹, A. Cano², M. Jiménez³

¹spereyra_c@hotmail.com

²Aromáticos La Victoria, SA de CV, San Luis Potosí, Mex. alicia_cano@arovic.com.mx

³Instituto de Ciencias Básicas. UV. maribjimenez@uv.mx

Resumen

En este estudio, se evaluó la liberación del sabor a través de la prueba sensorial tiempo-intensidad en gomas de mascar conteniendo microcápsulas elaboradas a base de diferentes materiales de pared (maltodextrina, concentrado de proteína de suero y pectina), en dicha prueba los panelistas determinaban la intensidad a menta con respecto al tiempo. Por otra parte, se analizó la textura de dicho producto tomando como parámetros importantes la dureza y adhesividad. Para la exploración de los datos se utilizó el análisis de varianza con intervalo de confianza de 0.05. Los resultados indicaron que la goma de mascar a base de microcápsulas de concentrado de proteína de suero-menta presentó mejores propiedades de textura y una intensidad máxima mayor respecto a las elaboradas con microcápsulas de pectina y maltodextrina. No obstante, la goma de mascar con microcápsulas de maltodextrina mostró una apariencia y grado de extinción menor que las otras dos gomas manejadas.

Introducción

La goma de mascar puede ser definida como un producto de dos fases, una insoluble en agua que a su vez está constituida por la goma base y la otra fase es soluble en agua la cual está compuesta por azúcares, edulcorantes, colorantes y saborizantes. Dicho producto es un buen modelo para estudiar la liberación del sabor, ya que ofrece la posibilidad de masticación de un producto semisólido por un periodo prolongado de tiempo, liberando progresivamente los compuestos de sabor. La velocidad con la que son liberados tales componentes depende principalmente del coeficiente de partición y la resistencia al transporte de masa, estos dos factores son afectados por la composición de los alimentos y textura, respectivamente (Ovejero et al., 2000).

Una alternativa para controlar la velocidad de liberación del sabor en productos alimenticios es el uso de la microencapsulación, proceso mediante el cual ciertas sustancias bioactivas son introducidas a una matriz con el fin de impedir su pérdida protegerlos de la reacción contra compuestos presentes en los alimentos o de factores externos (Madene et al., 2006). Existen diversas técnicas de encapsulación, en este estudio se utilizó el secado por aspersion que involucra la incorporación o dispersión del material a encapsular en el material de pared, seguido por la atomización de la solución dentro de una cámara de secado donde circula aire caliente y seco, las microcápsulas resultantes son transportadas al ciclón para su recolección (Ré, 1998). El sistema de pared de las microcápsulas está constituido por compuestos que generalmente contienen grupos hidrofóbicos o hidrofílicos, los cuales crean una estructura de red, tales como celulosas, gomas, lípidos, proteínas y

carbohidratos. Cabe mencionar, que la selección del material de pared dependerá de las propiedades del material a encapsular y las características deseadas del producto final.

El objetivo de este trabajo fue evaluar mediante la aplicación del método tiempo-intensidad la liberación del sabor en gomas de mascar, las cuales contenían microcápsulas de aceite esencial de menta a base de diferentes materiales de pared.

Material y métodos

Materia prima

La maltodextrina M100 (10DE) fue adquirida a través de Arancia S.A. (Ciudad de México), el concentrado de proteína de suero (80% proteína) fue provista por Vilher S.A. de C.V., la pectina de alto metoxilo (80% metoxilo) fue conseguida en La Nueva Serpentina S.A. de C.V. Por otra parte, respecto al aceite esencial de menta utilizado, fue donado por Aromáticos Victoria S.A. de C.V.

Preparación del sabor microencapsulado

Para la maltodextrina (MD) y concentrado de proteína de suero (CPS) se utilizó un contenido de sólidos del 30%, en el caso de la pectina (PEC) la concentración manejada fue del 5% debido a que a altas concentraciones de sólidos se generan soluciones de viscosidad alta. Estas soluciones se dejaron reposar durante toda la noche con la finalidad de garantizar la completa saturación de las moléculas. La relación aceite-sólidos fue de 1:4. La emulsión generada se mezcló con un agitador de aspas de alta velocidad (Cole Parmer Instrument Co.) a 5000 rpm por 20 min. Una vez preparadas las emulsiones, éstas fueron asperjadas en un secador por aspersion Büchi modelo 290 (Büchi-Laboratorios-Tecknick A.G., Flawil, Switzerland) con una temperatura de entrada de $200\pm 5^{\circ}\text{C}$ y una temperatura de salida de $110\pm 5^{\circ}\text{C}$.

Elaboración de las gomas de mascar

La formulación de la goma de mascar consistió en goma base, sorbitol, glucosa, glicerina, lecitina y microcápsulas. Como primer paso, se fundió la goma base a 80°C , posteriormente se fueron agregando los demás aditivos mezclando perfectamente con el fin de lograr una masa homogénea, hasta lo último y ya sin calentamiento se agregaron las microcápsulas de aceite esencial de menta, se mezcló y se moldeó para darle forma a la goma de mascar.

Análisis de textura

El estudio de la dureza y adhesividad de la goma de mascar, se realizó mediante pruebas de penetración, esto con la ayuda de un texturómetro TA-XT2i, (Texture Technologies Corp., Scarsdale, NY/Stable Microsystems, Godalming, UK).

Análisis sensorial

La prueba tiempo-intensidad fue realizada con 8 panelistas entrenados (3 hombres y 5 mujeres; con una rango de edad entre 18-23 años). El entrenamiento consistió en darles soluciones con diferentes concentraciones de menta, durante tres sesiones. Para obtener las curvas tiempo-intensidad, a los panelistas se les proporcionó una goma de mascar por día, ellos fueron marcando la intensidad percibida en diferentes lapsos de tiempo utilizando un escala no estructurada de 15cm. A los jueces se les indicó que la velocidad de masticación sería de una mordida por segundo. Por otra parte, a cada una de las curvas obtenidas se les

evaluó la intensidad máxima (Imax), el tiempo de máxima intensidad (Tmax), tiempo de extinción del sabor (Text), apariencia y grado de extinción.

Análisis de los datos

Para el exploración de los datos, se realizaron análisis de varianza y prueba de Tukey utilizando el paquete estadístico STATISCA (versión 7).

Resultados

Los análisis de textura realizados en las gomas de mascar indican que la GM-CPS (16.22N) presenta una dureza menor, seguida en orden ascendente por la GM-PEC (21.60N) y finalmente GM-MD (26.47N). La variación de la textura puede ser atribuida a la forma en que interactúan los materiales de pared en la goma de mascar, debido a que su composición solo varía en el tipo de microcápsulas añadidas. Cabe mencionar que se tomó una goma de referencia, ésta de marca comercial conocida (GM-REF), y se observó que la dureza obtenida por dicha goma (17.69N) es similar a la GM-CPS, realizando un análisis de varianza y una prueba de diferencia significativa honesta se determinó que estadísticamente no existen diferencias significativas entre la GM-CPS y GM-REF, pero si existen diferencias significativas ($p > 0.5$) entre GM-MD y GM-PEC con respecto a las otras dos gomas de mascar anteriormente mencionadas.

Tabla 1. Prueba de Tukey DSH (diferencia significativa honesta) para el parámetro dureza en las gomas de mascar.

| Gomas de mascar | X_{REF} (17.669) | X_{CPS} (16.229) | X_{MD} (26.471) | X_{PEC} (21.605) |
|-----------------|-----------------------|-----------------------|----------------------|-----------------------|
| GM-REF | | 0.279127 | 0.000233 | 0.003750 |
| GM-CPS | 0.279127 | | 0.000231 | 0.000616 |
| GM-MD | 0.000233 | 0.000231 | | 0.001033 |
| GM-PEC | 0.003750 | 0.000616 | 0.001033 | |

Existen diferencias significativas si $p < 0.05$.

Referente a la adhesividad existen diferencias significativas ($p > 0.5$) entre las cuatro diferentes gomas de mascar manejadas (Fig. 1), la goma de mascar más adhesiva fue la GM-CPS (-237.9N*s), seguida por GM-PEC (-166.66N*s), GM-MD (-6.35N*s) y por último GM-REF (-0.72N*s).

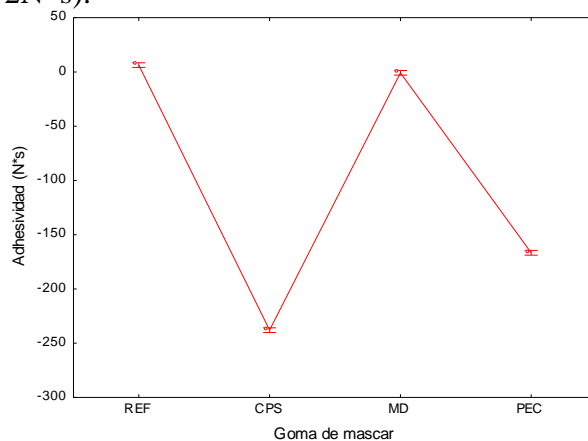


Fig. 1. Gráfico de medias para el parámetros adhesividad en las diferentes gomas de mascar utilizadas, las barras verticales denotan un intervalo de confianza del 95%.

Finalmente se evaluó la variación del sabor mediante las curvas tiempo intensidad que se muestran en la Fig. 2., como se observa la I_{max} mayor la presentó la GM-CPS a un T_{max} (144s) y la menor fue la registrada en la GM-PEC (23.2%) con un T_{max} (145s). La GM-MD mostró una I_{max} (26.3%) a un tiempo de 55 s. Respecto a la apariencia, se encontró que la GM-CPS posee una velocidad mayor para alcanzar la I_{max} , mientras que la GM-PEC obtuvo una tasa de incremento menor, probablemente debido a que la pectina dificulta la difusión de los componentes en el producto. Ahora bien, el grado de extinción que mostró la GM-MD fue menor al de la GM-CPS, no se pudo determinar este parámetro en la GM-PEC debido a que la pectina le confirió una textura desagradable a la goma de mascar y los panelistas desecharon la muestra antes de que terminara la prueba. El tiempo de extinción del sabor para la GM-CPS y GM-MD fue a los 10 minutos.

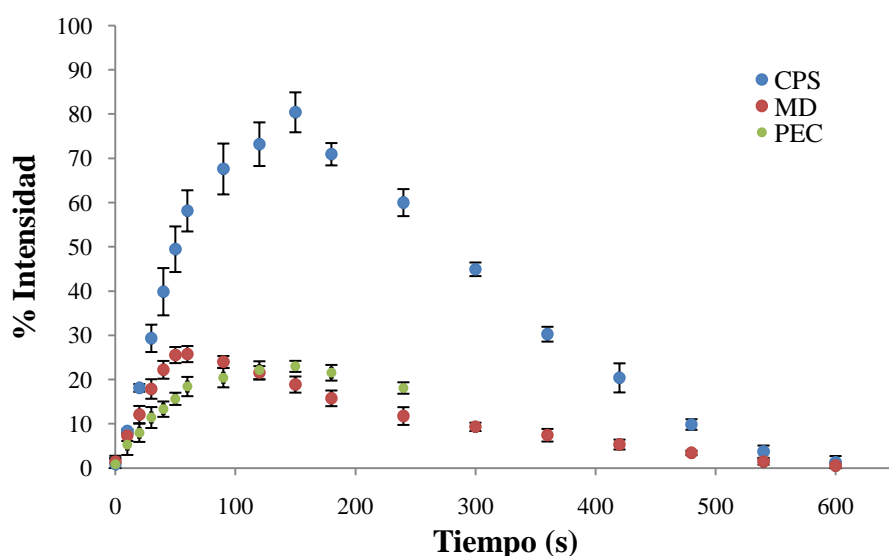


Fig. 2. Curva T-I promedio para las gomas de mascar elaboradas con microcápsulas a base de diferentes materiales de pared. Cada punto representa el promedio de las observaciones dadas por los panelistas.

Conclusiones

Estadísticamente no se encontraron diferencias significativas en cuanto al parámetro dureza entre la goma de mascar a base de microcápsulas de CPS y la goma de mascar de referencia. GM-PEC presentó una textura sensorialmente desagradable. En relación a la liberación del sabor en el producto, GM-CPS mostró la mayor I_{max} en comparación con las gomas de mascar en donde se utilizaron microcápsulas de PEC y MD. Por otra parte, probablemente exista un efecto de las microcápsulas sobre la difusión de los componentes en la goma de mascar, debido a que la apariencia y grado de extinción en las curvas T-I, varían respecto al material de pared utilizado para la elaboración del microencapsulado.

REFERENCIAS

- Baranauskienė, R., Bylaitė, E., Zukauskaitė, J., Venskutonis R.P. (2007). Flavor retention of peppermint (*Mentha piperita* L.) essential oil spray-dried in modified starches during encapsulation and storage. *Journal of Agricultural and Food Chemistry* 55, 3027-3036.
- Beristain C. I., Vásquez A., García H. S. y Vernon-Carter E. J. (1996). Encapsulation of orange peel oil by co-crystallization. *Lebensmittel-Wissenschaft und Technologie* 29, 645-647.
- Beristain, C.I., García, H.S. and Vernon, E.J. (1999). Mesquite gum (*Prosopis juliflora*) and maltodextrin blends as wall materials for spray-dried encapsulated orange peel oil. *Food Science and Technology International* 5, 353-356.
- Jafari S. M., Yinghe H. and Bhandari B. (2007). Encapsulation of nanoparticles of d-limonene by spray drying; Role of emulsifiers and emulsifying techniques. *Drying technology* 25, 1069-1079.
- Madene A., Jacquot M., Scher J., y Desobry S. (2006). Flavour encapsulation and controlled release – a review. *International journal food science and technology* 41, 1-21.
- Ovejero I., Haahr A., Van Der Berg F., y Bredie W. (2004). Flavor release measurement from gum model system. *Journal of Agricultural and Food Chemistry* 52, 8119-8126.
- Ré, M.I. (1998). Microencapsulation by spray drying. *Drying Technology* 16, 1195-1236.

Influencia de la Concentración de Almidón de Maíz y Mucilago de Nopal (*opuntia ficus indica*) en la Caracterización Mecánica de Películas Extrudidas Comestibles Utilizando el Análisis de Superficie de Respuesta

Morín-Alcázar C.S.¹, Calvo-López A.², Galicia-García T.³, Martínez-Bustos F.⁴, Grosso C.⁵, Aguilar-Palazuelos E.⁶

¹ Maestría en tecnología y ciencia de los Alimentos. UAQ. fanymorin@gmail.com

² CINESTAV-Unidad Querétaro

³ CINESTAV-Unidad Querétaro. tg Galicia@gro.cinvestav.mx

⁴ CINESTAV-Unidad Querétaro. fmartinez@gro.cinvestav.mx

⁵ Universidad Estatal de Campinas. Brasil. grosso@fea.unicamp.br

⁶ Maestría en Ciencia y Tecnología de los Alimentos. Universidad Autónoma de Sinaloa. eaguilar@uas.uasnet.mx

La creciente actividad en los procesos de maduración en frutas hace necesaria la búsqueda y desarrollo de métodos de conservación a base de fuentes naturales que garanticen la inocuidad e integridad hasta su consumo. En el presente trabajo se cuantificó la influencia que presenta la concentración de almidón de maíz ceroso (60-80%), mucilago de nopal (10-30%) y plastificante (10-30%) en la elaboración de películas comestibles para el recubrimiento de frutos. Las mezclas fueron procesadas en un extrusor compacto (CINESTAV, Querétaro) y las películas obtenidas se caracterizaron en sus principales propiedades mecánicas de resistencia máxima a la tracción (σ_{\max}) y tensión a la ruptura (ϵ) (ASTM-D882-00, 2001). El análisis de los resultados se realizó de acuerdo a un diseño de mezclas con tres variables, a través del paquete estadístico Design Expert v6.010 (Stat-Ease, 2003), obteniendo un número de 11 unidades experimentales y el comportamiento de las variables significativas se realizó mediante la interpretación de datos obtenidos mediante el análisis de superficies de respuesta. Se concluyó que: 1) La variable σ_{\max} fue afectada por la interacción almidón mucilago A-M ($p < 0.1$). 2) La ϵ fue afectada por las interacciones almidón-plastificante A-P ($p < 0.0192$), y plastificante-mucilago ($p < 0.0223$). 3) Las propiedades mecánicas de σ_{\max} y ϵ fueron favorecidas a concentraciones intermedias de almidón ceroso (70%), plastificante y mucilago (20%), en donde la película pudo ser traccionada hasta un 30% de su espesor original.

Palabras clave: película comestible, almidón ceroso, mucilago de nopal propiedades mecánicas, extrusión

AGRADECIMIENTOS: CONACYT, CONCYTEQ por el apoyo económico brindado para la realización de este proyecto.

Introducción

Las películas plásticas comestibles son materiales que solidifican en condiciones ambientales y están conformadas por polímeros de origen glucosídico, compuestos plastificantes de bajo peso molecular (polioles) y aditivos (pectinas) que pueden ser aplicadas como membranas protectoras, de una manera similar a como lo hiciera la naturaleza con las epidermis de los frutos, permitiéndole una conservación prolongada de alimentos preparados y, en el caso de frutas y hortalizas, un control en los procesos de

maduración (Arizmendi, 2004). El almidón de maíz así como el mucílago de nopal poseen potencialmente ciertas propiedades en su composición, que procesados mediante el proceso termo-mecánico de extrusión se inducen modificaciones físicas y químicas que favorecen su plastificación (Olkku y Linko, 1977) y la elaboración de películas comestibles y biopelículas. A través del diseño de mezclas es posible el análisis y la obtención de los materiales con mejores proporciones en base a las concentraciones utilizadas en sus ingredientes. El objetivo de este trabajo fue la elaboración y caracterización mecánica de películas comestibles a base de almidón de maíz ceroso y mucílago de nopal *Opuntia ficus* usando el proceso de extrusión para la elección de los mejores composiciones.

Condiciones experimentales

Materias primas y acondicionamiento de las muestras.

Se utilizó almidón de maíz waxy nativo grado alimenticio (Almex, México, D.F), glicerol como plastificante (Sigma-Aldrich, Alemania) y mucílago de nopal (*Opuntia ficus indica*) extraído de acuerdo al método reportado por Arizmendi (2004). Las proporciones de cada ingrediente se realizaron de acuerdo a un diseño de mezclas (Bowles y Montgomery, 1997). Una vez obtenida la mezcla se refrigeró a 5°C por 24h.

Procesamiento y caracterización de la muestra.

La elaboración de las películas usando la tecnología de extrusión termoplástica se basó en la metodología descrita por Galicia-García y col., (2007). Se utilizó un extrusor de tornillo simple marca CINVESTAV (Querétaro, México) con tres zonas de calentamiento (60, 110 y 75 °C, en las zonas de alimentación, transición y compresión respectivamente), y tornillo con relación de compresión 1:1. Todas las mezclas fueron alimentadas al extrusor a una velocidad de 18 rpm. Se utilizó una matriz rectangular 40 mm de ancho y 0.75 mm de espesor. Las películas obtenidas fueron caracterizadas en sus propiedades mecánicas de resistencia máxima a la tracción ($\sigma_{\text{máx}}$) y alargamiento a la ruptura (ϵ), de acuerdo con la norma (ASTM D882-00, 2001), siendo la distancia inicial de separación fija en 50mm con una velocidad de realización de la prueba en 1.0mm/s. Las probetas obtenidas de las películas (100 X 25mm) fueron analizadas en un texturómetro TA.XT2 (Stable Micro System, Surrey, UK), con 6 repeticiones para cada muestra (Figura 1).



Figura1. Texturómetro TA.XT2

Diseño y análisis de los datos.

La preparación de las mezclas para la elaboración de las películas comestibles, fue realizada siguiendo un diseño experimental de mezclas, de forma aleatoria con una replica, donde se utilizó una variación en la concentración para el almidón ceroso de (60- 80%),

mucílago de nopal (10-30%) y para el glicerol como plastificante (10-30%). En la Tabla 1 se presenta el diseño de 3 componentes con sus respectivos niveles. La significancia se realizó mediante un análisis de varianza (prueba F) y el efecto de las mezclas fue llevado a cabo aplicando la metodología de superficie de respuesta mediante el paquete estadístico Design Expert (State-Ease, 2003).

Tabla 1. Diseño experimental de mezclas, para la elaboración de las películas comestibles.

| Tratamiento | Almidón ceroso (X1) | Glicerol (X2) | Mucílago (X3) |
|-------------|---------------------|---------------|---------------|
| 1 | 0.80 | 0.10 | 0.10 |
| 2 | 0.67 | 0.17 | 0.17 |
| 3 | 0.70 | 0.20 | 0.10 |
| 4 | 0.60 | 0.20 | 0.20 |
| 5 | 0.63 | 0.23 | 0.13 |
| 6 | 0.73 | 0.13 | 0.13 |
| 7 | 0.70 | 0.10 | 0.20 |
| 8 | 0.60 | 0.30 | 0.10 |
| 9 | 0.80 | 0.10 | 0.10 |
| 10 | 0.63 | 0.13 | 0.13 |

Resultados y Discusión

Resistencia Máxima a la Tracción ($\sigma_{\text{máx}}$). Las películas obtenidas por el proceso de extrusión presentaron valores experimentales en un rango de 1.559 a 10.431 MPa. El análisis de regresión mostró que la $\sigma_{\text{máx}}$ fue afectada por la interacción concentración de almidón- mucílago A-M ($p \leq 0.10$). El modelo de predicción para el $\sigma_{\text{máx}}$ utilizó las variables codificadas:

$$Y \sigma_{\text{máx}} = +7.75 * X1 + 3.34 * X2 + 1.07 * X3 - 10.48 * X1 * X2 + 16.81 * X1 * X3$$

El modelo de predicción utilizado explicó el 69.51% de la variación total ($p \geq 0.10$), para los valores de resistencia máxima la tracción ($\sigma_{\text{máx}}$). De manera general se observó que al disminuir el contenido de mucílago (30 a 10 %) se presentó un incremento en los valores de $\delta_{\text{máx}}$ (Figura 2b), y a valores intermedio del plastificante (20%) se presentaron valores altos de $\delta_{\text{máx}}$, este comportamiento fue debido principalmente a la incorporación de grupos OH que incrementaron los enlaces de hidrógeno presentes (Van Soest y Knooren, 1997).

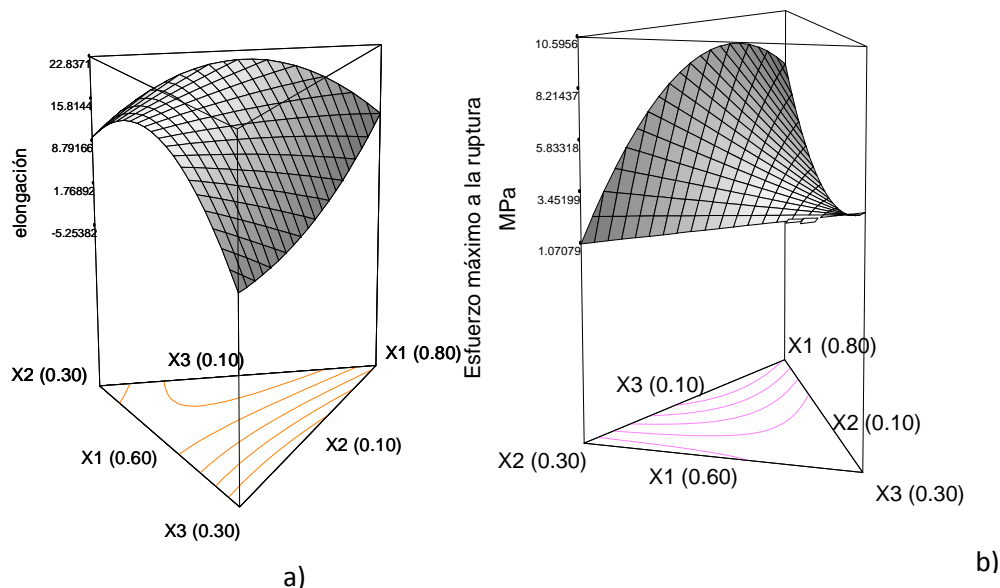


Figura 2. Gráficos representativos de las propiedades mecánicas evaluadas: (a) gráfico de interacción almidón-plastificante- mucílago en ϵ y, (b) gráfico de interacción almidón-plastificante- mucílago en $\delta_{\text{máx}}$.

Alargamiento a la ruptura (ϵ). Se presentaron valores experimentales en un rango de 3.756 a 19.709 %. El análisis de regresión mostró que ϵ fue influenciada por la interacción de almidón y de plastificante A-P ($p \leq 0.0192$). El modelo de predicción para el ϵ utilizó las variables codificadas:

$$Y_{\epsilon} = 11.29 * X1 + 9.66 * X2 - 0.63 * X3 + 43.45 * X1 * X2 - 15.19 * X1 * X3$$

El modelo de predicción utilizado explicó el 90.33% de la variación total ($p \leq 0.05$) para los valores de elongación. De manera general se observó que a concentraciones de almidón ceroso intermedio (70%) y valores de glicerol y de mucílago de (10 a 20%) se favorecieron los valores de ϵ (Figura 2a), esto debido a la combinación de éstos ya que al incorporarse a la estructura del almidón favorecen su flexibilidad (Arizmendi, 2004).

Conclusiones. 1) La respuesta $\sigma_{\text{máx}}$ fue afectada por la interacción concentración de almidón- mucílago A-M ($p \geq 0.10$); a bajas concentraciones de mucílago (10 y 13%), se incrementó $\sigma_{\text{máx}}$. 2) ϵ fue afectado por la interacción de almidón- plastificante A-P ($p \geq 0.0192$), el valor de ϵ decreció a medida que se incrementó el contenido de almidón A (70 a 80%), el contenido de mucílago y plastificante (20 a 30%) siendo favorecidos en concentraciones donde el almidón varió de 70 a 73% y mucilago y plastificante de 10 a 20%. 3) Mediante el análisis de los gráficos de superficie fue posible el observar el efecto de cada una de las mezclas en la elaboración de películas comestibles.

REFERENCIAS

- Arizmendi Coter Daniel 2004. Optimización de dos compuestos plastificantes (glicerol y polietilenglicol) en la elaboración de una película plástica comestible obtenida a partir del mucílago de nopal de la especie *Opuntia tomentosa* Salm-Dyck. Tesis de Maestría. Universidad Autónoma de México.
- American Society for Testing and Materials (ASTM) 2001. Standard Test Methods for Tensile Properties of Thin Plastic Sheeting. ASTM D882-00. In: ASTM Standards on Disc. Philadelphia, ASTM v.08.01. 9p., 2001.
- Bowles, Melissa L., and Montgomery, Douglas C. 1997. How to formulate the ultimate margarita: a tutorial on experiments with mixtures. *Quality Engineering*, 10(2):239-253.
- Galicia-García T., Martínez-Bustos F., Jiménez-Arévalo O., and Aguilar-Palazuelos E. 2007. Mechanical Properties of Extruded Biodegradable films of native starch from sugar cane fibre. *Food Engineering Research Developments*. Chapter 9. ISBN:978-1-60021-906-1
- Olkku J y Linko P. 1977. Effects of thermal processing on cereal based food systems. En: *Food quality and nutrition research priorities for thermal processing*. Dawney, EK. Eds. London. Applied Science Publishers. Pp. 352.
- State-Ease. 2003. *Design-Expert*. V 6.010. East 2021 Hennepin Av. Minneapolis, Mn 55413.
- Van Soest, J; and Knooren. 1997. Influence of glycerol and water content on the structure and properties of extruded starch plastics sheets during aging. *J.Appl. Polym. Sci.* 64: 1411- 1422

Efecto de la Concentración de Almidón de Haba (vicia faba l) en las Propiedades Mecánicas de Películas Comestibles.

Calvo-López A.¹, Morín-Alcázar C.S.², Galicia-García T.³, Martínez-Bustos F.⁴

¹CINVESTAV-Unidad Querétaro. camd_86@hotmail.com

²Maestría en tecnología y ciencia de los Alimentos, Universidad Autónoma de Querétaro.

fmartinez@gro.cinvestav.mx

³CINVESTAV-Unidad Querétaro. grosso@fea.unicamp.br

⁴CINVESTAV-Unidad Querétaro. tg Galicia@gro.cinvestav.mx

Las películas o biopelículas son materiales elaborados a partir de materiales comestibles que pueden actuar como barrera ante elementos externos y consecuentemente, proteger y disminuir el deterioro del alimento recubierto, además pueden proveer nutrientes, u otros aditivos funcionales. El presente trabajo se estableció con la finalidad de evaluar el efecto del almidón de haba en diferentes concentraciones (60-80%, en la formulación) como soporte principal de la matriz polimérica en la elaboración de películas comestibles. También se utilizaron concentraciones establecidas (10-30%) de plastificante y mucílago de nopal. Las mezclas fueron procesadas en un extrusor compacto (CINVESTAV, Querétaro) y las películas obtenidas fueron caracterizadas en sus propiedades mecánicas de resistencia máxima a la tracción ($\sigma_{\text{máx}}$), alargamiento a la ruptura (ϵ) y módulo de elasticidad (E) de acuerdo a norma (ASTM-D882-00, 2001). El análisis de los resultados se realizó utilizando un diseño de mezclas con tres variables (fuente de almidón, plastificante y aditivo), a través del paquete estadístico Design Expert v6.010 (Stat-Ease, 2003). El comportamiento de las variables significativas se analizó en base a la tendencia presentada en gráficos de superficie de respuesta. En el presente trabajo se concluyó: que la concentración de almidón de haba influye directamente sobre las propiedades mecánicas evaluadas (ϵ , E y $\sigma_{\text{máx}}$), siendo favorecidas a concentraciones intermedias (70%) de almidón de haba en películas obtenidas por el proceso de extrusión termoplástica.

Palabras clave: Película comestible, Almidón de haba, metodología de superficie de respuesta, extrusión.

AGRADECIMIENTOS: CINVESTAV, por las facilidades otorgadas para la realización de este proyecto.

Introducción

La funcionalidad de polímeros orgánicos ha emergido recientemente como parte importante de la ciencia y la tecnología, principalmente en la elaboración de materiales de fuentes naturales. El almidón se ha considerado durante muchos años como un polímero con un alto potencial para la fabricación de recubrimientos en alimentos debido a que es una fuente de bajo costo, renovable y biodegradable (Aminabhavi, 1990). El aprovechamiento de los recursos que se generan en el país hace que las materias primas tradicionales como es el caso de almidón de haba y mucílago de nopal representen una alternativa en el uso del mejoramiento de las propiedades de películas comestibles. El método mas utilizado en la elaboración de recubrimientos es a través de casting por vaciado (Bello-Pérez, *et al.*,2002).

Aunque, debido a la alta viabilidad que presenta el almidón al ser procesado en equipos utilizados comúnmente para plásticos, hace que mediante el uso de técnicas como la extrusión sea posible la obtención de películas comestibles de aplicación directa al alimento, con alta inocuidad (Brown, 1981). El presente trabajo tuvo como objetivo la evaluación de diferentes concentraciones de almidón de haba en la elaboración de películas comestibles a base de almidón de haba, utilizando como aditivo mucílago de nopal.

Condiciones experimentales.

Materias primas y acondicionamiento de las muestras

Se utilizó como materia prima, almidón de haba (*Vicia faba L*), mucílago de nopal (*Opuntia ficus indica*) como aditivo y glicerol como plastificante (Sigma Aldrich, Alemania). La extracción del almidón de haba se realizó en base a metodología reportada por Escobar (2007) y la extracción de mucílago en base al método reportado por Arizmendi (2004). Las proporciones de cada ingrediente se realizaron de acuerdo a un diseño de mezclas (Bowles y Montgomery, 1997). Una vez obtenida la mezcla se refrigeró a 5°C por 24h.

Procesamiento y caracterización de la muestra.

La elaboración de las películas por la tecnología de extrusión termoplástica se basó en la metodología descrita por Galicia-García et al (2007). Un extrusor de tornillo único marca CINVESTAV (Querétaro, México) fue utilizado para la elaboración de las películas comestibles. Se mantuvieron constantes las zonas de alimentación, transición y compresión (60, 110 y 75 °C), el tornillo utilizado fue con una relación de compresión de 1:1. El formato de la película se realizó usando una matriz rectangular (40 mm X 0.75 mm).

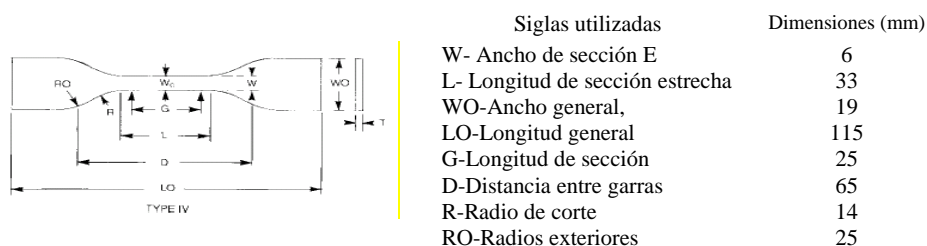


Figura1. Dimensiones utilizadas en probeta tipo IV, para la caracterización mecánica en películas comestibles.

Las películas obtenidas fueron caracterizadas en sus propiedades mecánicas (resistencia máxima a la tracción $\sigma_{\text{máx}}$, alargamiento a la ruptura ϵ y módulo de elasticidad E), de acuerdo con la norma ASTM D882-00, 2001. Las probetas obtenidas de las películas fueron tipo IV en base a norma (Figura 1). Los ensayos se analizaron en un texturómetro TA.XT2 (Stable Micro System, Surrey, UK), con 3 repeticiones para cada muestra.

Diseño y análisis de los datos.

Se utilizó un diseño experimental de mezclas, con 3 componentes (X1=Almidón, X2=Plastificante, X3=mucílago) (Bowles y Montgomery, 1997) y una replica. Las concentraciones utilizadas de almidón de haba fueron de 60 a 80%, mientras que para el mucílago de nopal y glicerol de 10 a 30% (Tabla 1). El efecto de las mezclas se realizó a

través de metodología de superficie de respuesta y la significancia del modelo mediante un análisis de varianza (prueba F) en el paquete estadístico Design Expert (State-Ease, 2003).

Tabla 1. Diseño experimental de mezclas, para la elaboración de las películas comestibles a base de almidón de haba.

| Tratamiento | Almidón (X1) % | Glicerol (X2) % | Mucílago (X3) % |
|-------------|----------------------|-----------------------|-----------------------|
| 1 | 80 | 10 | 10 |
| 2 | 67 | 17 | 17 |
| 3 | 70 | 20 | 10 |
| 4 | 60 | 20 | 20 |
| 5 | 63 | 23 | 13 |
| 6 | 73 | 13 | 13 |
| 7 | 70 | 10 | 20 |
| 8 | 60 | 30 | 10 |
| 9 | 80 | 10 | 10 |
| 10 | 60 | 10 | 30 |
| 11 | 63 | 13 | 23 |

Resultados y Discusión.

Alargamiento a la ruptura (ϵ). Se presentaron valores experimentales en un rango de 15.9 a 42.59 %. El análisis de regresión mostró que el ϵ fue influenciada por la interacción del almidón y el plastificante A-P ($p \leq 0.1$) y almidón de mucílago ($p \leq 0.1$). El modelo de predicción para el ϵ utilizó las variables codificadas:

$$\epsilon = 0.90 * X_1 + 15.89 * X_2 + 22.19 * X_3 + 74.74 * X_1 * X_2 + 77.06 * X_1 * X_3$$

El modelo de predicción utilizado explicó el 69.95% de la variación total ($p \leq 0.1$). De manera general se observó que a concentraciones de 15 a 20 % de plastificante y mucílago, así como 70% de almidón de haba, se favorecieron los valores de ϵ (Figura 2c). Este comportamiento pudo ser debido a la interacción presente entre el plastificante y el almidón (mediante enlaces de hidrógeno) que en conjunto con el mucílago le proporcionaron a la película obtenida un mayor alargamiento al ser traccionada durante el proceso de extrusión (Guilbert, 1986).

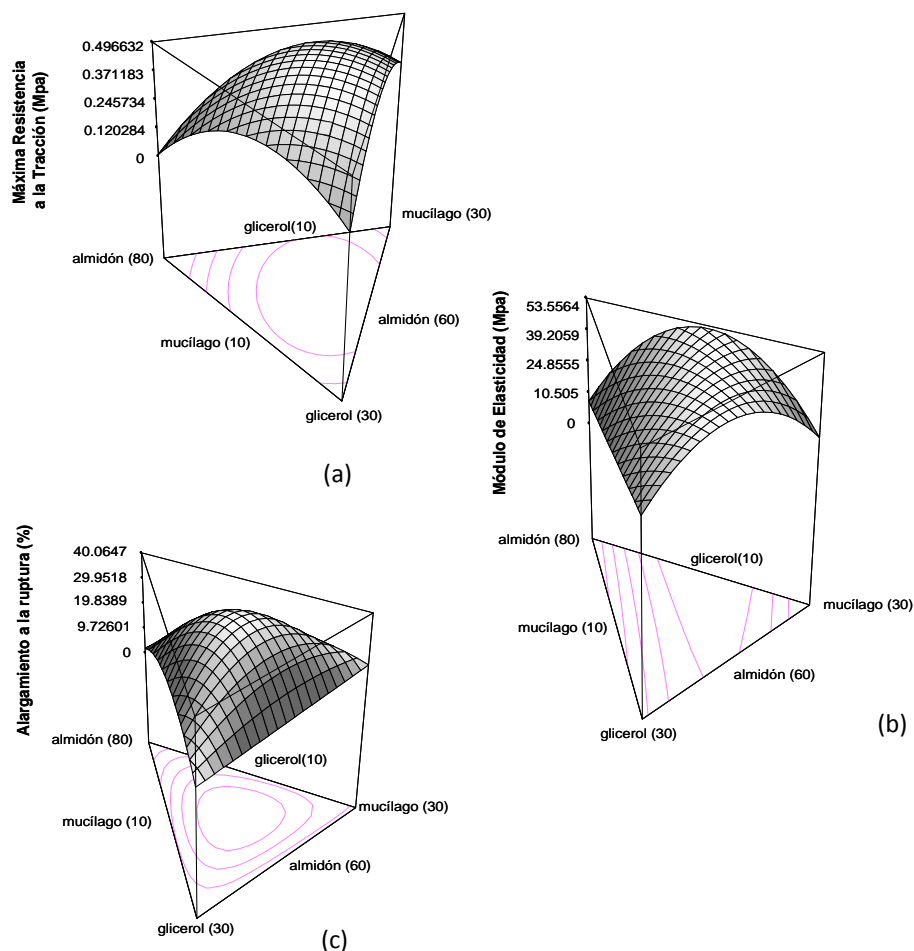


Figura 2. Gráficos representativos de las propiedades mecánicas evaluadas: (a) interacción almidón- mucílago en $\sigma_{\text{máx}}$, (b) interacción almidón- mucílago en E (c) interacción almidón-plasticante en ϵ .

Módulo de Elasticidad (E). Los valores experimentales se presentaron en un rango de 18.53 a 58.69 MPa. El análisis de regresión mostró que E fue influenciada por la interacción almidón-mucílago A-M ($p \leq 0.1$). El modelo de predicción para el ϵ utilizó las variables codificadas:

$$E = 6.74 * X_1 + 27.14 * X_2 + 17.21 * X_3 + 161.10 * X_1 * X_3 + 93.16 * X_2 * X_3$$

El 59.52% de la variación total ($p \leq 0.1$) fue explicada a través del modelo de predicción. De manera general se observó que a concentraciones del 15% de plasticante, 10% de mucílago y 70% de almidón de haba se presentó un decremento en los valores de E, indicando que a estas concentraciones las películas obtenidas presentan una mayor flexibilidad, a diferencia de cuando se disminuye el contenido de almidón (60%) en donde las películas presentaron mayor rigidez (Lourdin *et al.*, 1995).

Resistencia máxima a la tracción ($\sigma_{\text{máx}}$). Las películas comestibles elaboradas por el proceso de extrusión termoplástica presentaron valores de 0.30 a 0.48 MPa en este parámetro. El análisis de regresión mostró que la $\sigma_{\text{máx}}$ fue afectada por la interacción de

almidón-plastificante A-P ($p \leq 0.39$), almidón-mucílago A-M ($p \leq 0.0025$) y plastificante-mucílago P-M ($p \leq 0.0120$). El modelo de predicción para la σ_{\max} utilizó las variables codificadas:

$$\sigma_{\max} = 2.602E-03 * X_1 + 0.27 * X_2 + 0.28 * X_3 + 0.93 * X_1 * X_2 + 1.03 * X_1 * X_3 + 0.74 * X_2 * X_3$$

El modelo de predicción utilizado explicó el 96.88% de la variación total ($p \leq 0.0009$), para los valores de resistencia máxima de tracción. De manera general se presentaron altos valores de σ_{\max} a valores intermedios de almidón (70%), mucílago y plastificante (20%), en comparación con altas concentraciones de cada uno de los componentes de la mezcla donde se presentaron bajos valores de σ_{\max} y una película con menor resistencia.

Conclusiones. 1) La respuesta ε fue afectada por la interacción concentración de almidón-plastificante A-P ($p \geq 0.10$) y almidón mucílago A-M ($p \geq 0.10$). 2) La variable respuesta E, fue afectada por la interacción almidón mucílago A-M ($p \geq 0.10$). 3) Lo que respecta a Máxima resistencia a la tracción σ_{\max} , la variación de este parámetro fue afectada por la interacción de almidón-plastificante A-P ($p \leq 0.39$), almidón-mucílago A-M ($p \leq 0.0025$) y plastificante-mucílago P-M ($p \leq 0.0120$). 4). A través del análisis de los datos mediante los gráficos de superficie de respuesta fue posible observar que la concentración de almidón influye directamente en las propiedades mecánicas, obteniéndose mejores propiedades a valores intermedios (70%). 5) La incorporación de plastificante y mucílago a concentraciones menores de 20% si favorecieron las variables respuesta analizadas.

Bibliografía.

- Aminabhavi, T. M., Balundgi, R. H., and Cassidy, P.E. 1990. A review on biodegradable plastics. *Polymer Plastics technology and Engineering*, 29:235-262.
- Arizmendi Coteró Daniel 2004. Optimización de dos compuestos plastificantes (glicerol y polietilenglicol) en la elaboración de una película plástica comestible obtenida a partir del mucílago de nopal de la especie *Opuntia tomentosa* Salm-Dyck. Tesis de Maestría. Universidad Autónoma de México.
- American Society for Testing and Materials (ASTM) 2001. Standard Test Methods for Tensile Properties of Thin Plastic Sheeting. ASTM D882-00. In: ASTM Standards on Disc. Philadelphia, ASTM v.08.01. 9p., 2001.
- Bello-Pérez A., Contreras-Ramos M., Romero-Manilla R., Solorza-Feria J y Jiménez-Aparicio A. 2002. Chemical and functional properties of modified starches from banana (*Musa paradisiaca*) Var. Macho. *Agrociencia* 36:169-180.
- Bowles, Melissa L., and Montgomery, Douglas C. 1997. How to formulate the ultimate margarita: a tutorial on experiments with mixtures. *Quality Engineering*, 10(2):239-253.
- Brown W.E. 1981. Handbook of Plastics test methods. 2nd ed. George Godwin. London. p423.
- Galicia-García T., Martínez-Bustos F., Jiménez-Arévalo O., and Aguilar-Palazuelos E. 2007. Mechanical Properties of Extruded Biodegradable films of native starch from sugar cane fibre. *Food Engineering Research Developments*. Chapter 9. ISBN:978-1-60021-906-1
- Stat-Ease 2001. Design-expert version 6.0.5. Stat-Ease, Inc. East 2021 Hennepin Ave., Suite 1991. Minneapolis, MN 55413.

Propiedades Mecánicas en Láminas para Bolsas Biodegradables Obtenidas por el Proceso de Extrusión Termoplástica. Análisis de Superficie de Respuesta

Galicia-García, T.¹, Martínez-Bustos, F.², Jiménez-Arévalo, O.³

¹ CINEVESTAV-Unidad Querétaro. tg Galicia@gro.cinvestav.mx

² CINEVESTAV-Unidad Querétaro. fmartinez@gro.cinvestav.mx

³ CIATEQ. omar.jimenez@ciateq.mx

Ante el creciente aumento de residuos sólidos generados en el país, el uso de materiales provenientes de fuentes naturales para la elaboración y desarrollo de sustitutos de plásticos ha sido de gran interés, por lo que es necesario durante su obtención el análisis de cada una de las variables de proceso utilizadas a través de la construcción de diseños experimentales compactos que ayuden al manejo e interpretación de los datos obtenidos, como es el caso de la metodología de superficie de respuesta. En el presente trabajo se estudió el efecto del contenido de humedad H (18.25- 22.75 %), fibra F (0-10%) y modificación realizada al almidón de maíz ceroso A (fosfatación y acetilación) en láminas obtenidas por el proceso de extrusión termoplástica. Las mezclas fueron procesadas en un extrusor compacto (CINEVESTAV, Querétaro) y las láminas obtenidas fueron caracterizadas en sus principales propiedades mecánicas (resistencia máxima a la tracción σ_{\max} , alargamiento a la ruptura ϵ y módulo de elasticidad E) de acuerdo a norma (ASTM-D882-00, 2001). Los resultados obtenidos fueron analizados de acuerdo con un Diseño Central compuesto con 2 factores numéricos (contenidos de humedad y fibra) y 1 categórico (Modificación de almidón) mediante el paquete estadístico Design Expert v6.010 (Stat-Ease, 2003). Fue establecido un mínimo de puntos centrales y replica del punto factorial, dando así un total de 15 unidades experimentales. La variable respuesta en donde el modelo no fue significativo los datos fueron analizados mediante gráficos de interacción y de un solo factor. De manera general se concluyó que las tendencias observadas en los gráficos obtenidos facilitaron el manejo e interpretación de las propiedades mecánicas de las láminas obtenidas por el proceso de extrusión.

Palabras clave: Almidón termoplástico, metodología de superficie de respuesta, análisis de interacción y de un solo factor.

AGRADECIMIENTOS: CONACYT, CONCYTEQ por el apoyo económico brindado para la realización de este proyecto.

Introducción

A pesar de los usos que tiene el almidón en su estado nativo, su eficiencia disminuye al someterse a distintas condiciones de proceso como temperatura y presión, por lo que a través de su modificación química, como es el caso de la fosfatación y acetilación, es posible proporcionar al almidón propiedades funcionales a la estructura nativa de los principales componentes del almidón (amilosa y amilopectina) (Landerito y Wang., 2005). Los materiales como láminas y películas obtenidos en base a biopolímeros elaborados principalmente de diversas fuentes como polisacáridos, en donde algunas características intrínsecas de láminas (punto de fusión, hidrofiliidad, compatibilidad con los ingredientes,

entre otros) hacen que su procesamiento por extrusión sea más complejo que el obtenido por otros métodos comúnmente empleados (casting por vaciado) (Cho *et al.*, 2002) y con un pocos estudios publicados (Park *et al.*, 2008). Uno de los inconvenientes durante la obtención del material es el análisis de la influencia de cada una de las variables del proceso en las respuestas obtenidas, por lo cual es indispensable hacer uso de herramientas estadísticas que nos ayuden a restringir las variables, los niveles de proceso y el análisis de sus variables respuesta antes y después de la obtención de las láminas biodegradables, como es el caso de un diseño central compuesto y análisis del comportamiento de los datos a través de la metodología de superficie de respuesta.

Condiciones Experimentales

Materias primas y acondicionamiento de la muestra.

Se utilizó almidón de maíz waxy nativos grado alimenticio (Almex, México D.F), glicerol como plastificante (Sigma-Aldrich, Alemania), y bagazo de caña de azúcar previamente tratada (Zafra 2005, Veracruz, México) como fuente de fibra. La modificación del almidón se realizó de acuerdo a la metodología reportada por San Martín-Martínez *et al.*, 2004 (fosfatación), y Bello-Pérez *et al.* (2002) (acetilación). Las muestras se mezclaron y acondicionaron en una proporción y a una humedad establecida de acuerdo al diseño central compuesto. La mezcla obtenida se pasó a través de una malla de 0.5 mm y se conservó en cámara de refrigeración a 5 °C por 12 horas hasta su posterior uso.

Procesamiento y caracterización de la muestra

Las mezclas acondicionadas fueron procesadas en un extrusor monorosca prototipo de laboratorio marca CINVESTAV, Querétaro, México, con una relación de compresión 1:1, un dado con medidas internas de 40 mm de ancho por 0.75 mm de espesor. La temperatura en la zona de transición (zona 2) fue establecida de acuerdo con el diseño experimental entre 110 y 140 °C, la temperatura de las zonas 1 y 3 permanecieron constantes a 60 y 75 °C respectivamente. El resto de los parámetros de extrusión fueron constantes de acuerdo a ensayos preliminares (Galicia-García *et al.*, 2007). Una vez obtenidas las láminas fueron caracterizadas en sus principales propiedades mecánicas (resistencia máxima a la tracción $\sigma_{m\acute{a}x}$, alargamiento a la ruptura ϵ y modulo de elasticidad E) de acuerdo a norma técnica (ASTM D882-00, 2001).

Diseño y análisis de los datos

A través de información bibliográfica y ensayos preliminares fueron determinados los niveles de las variables independientes, así como su codificación (X1=H=Humedad de acondicionamiento %, X2=F=Fibra %, X3=A=Tipo de Almidón. Otras condiciones que se mantuvieron constantes fueron las dimensiones del dado del extrusor (40mm X0.75 mm), tamaño de partícula (<0.5mm), velocidad del tornillo y de alimentación (30 y 25 rpm). La matriz de diseño utilizada para las muestras de almidón y maíz ceroso (Tabla 1), fueron analizados de acuerdo con un Diseño Central compuesto con 2 factores numéricos (contenidos de humedad y fibra) y 1 categórico (Modificación de almidón) con 3 niveles (almidón fosfatado, acetilado y normal). Fue establecido un mínimo de puntos centrales y replica del punto factorial, dando así un total de 15 unidades experimentales. La repetición en el punto central permitió la estimación de la variabilidad. Los datos experimentales se utilizaron para determinar los coeficientes ($\beta_{\square 0}$) de la ecuación de expansión de la serie de

Taylor (Polinomial cuadrática). La significancia del modelo se justificó mediante el análisis de la varianza (prueba F), y la influencia de las variables fueron obtenidas a través de los gráficos de superficie de respuesta. La variable respuesta en donde el modelo no fue significativo los datos fueron analizados mediante gráficos de interacción y de un solo factor.

Tabla 1 Matriz de diseño utilizado^a en láminas a base de almidón de maíz ceroso.

| Ensayo | Variables de Proceso | | |
|--------|----------------------|--------|-------|
| | H (X1) | F (X2) | A(X3) |
| 1 | 18.25 | 0 | Fosf |
| 2 | 22.75 | 0 | Fosf |
| 3 | 18.25 | 10 | Fosf |
| 4 | 22.75 | 10 | Fosf |
| 5 | 20.5 | 5 | Fosf |
| 6 | 18.25 | 0 | Act |
| 7 | 22.75 | 0 | Act |
| 8 | 18.25 | 10 | Act |
| 9 | 22.75 | 10 | Act |
| 10 | 20.5 | 5 | Act |
| 11 | 18.25 | 0 | Norm |
| 12 | 22.75 | 0 | Norm |
| 13 | 18.25 | 10 | Norm |
| 14 | 22.75 | 10 | Norm |
| 15 | 20.5 | 5 | Norm |

H: humedad, F: fibra, A: tipo de almidón:fosf(fosfatado), act (acetilado), Norm(nativo)

Resultados y discusión

Resistencia Máxima a la Tracción ($\sigma_{\text{máx}}$): Láminas obtenidas por el proceso de extrusión presentaron valores experimentales en un rango de 1.004 a 3.151 MPa. El análisis de regresión mostró que la $\sigma_{\text{máx}}$ fue afectada por la interacción concentración de humedad-fibra H-F ($p \leq 0.10$). Debido a que el modelo no fue significativo ($p \leq 0.1$), para la interpretación de las variables se realizó el análisis de acuerdo a gráfico de interacción (Figura 1a) representativo, utilizando el paquete estadístico Desig Expert v6.010 (Stat-Ease, 2003). De manera general se observó que al incrementarse la humedad (20.50 a 22.75 %) se presentó un decremento en los valores de $\delta_{\text{máx}}$, siendo más visible este comportamiento en almidones modificados, debido principalmente a la introducción de grupos hidrofílicos que favorecieron la formación de enlaces de hidrógeno (Miles *et al.*, 1985), presentándose en menor proporción en lámina a base de almidón nativo debido a un mayor orden estructural presente en la estructura ramificada de la amilopectina (López *et al.*, 2008).

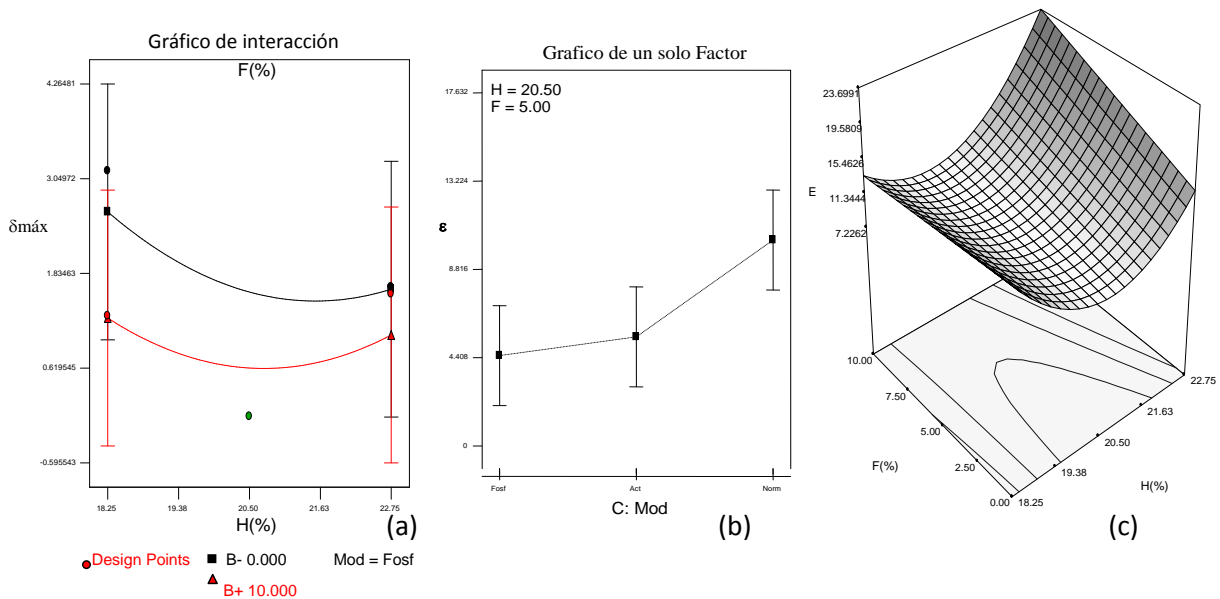


Figura 1. Gráficos representativos de las propiedades mecánicas evaluadas: (a) gráfico de interacción humedad-fibra en $\delta_{máx}$ de almidón fosfatado, (b) gráfico de un solo factor en ϵ de almidón acetilado, (c) gráfico de superficie de respuesta en E de almidón nativo.

Alargamiento a la Ruptura (ϵ). Se presentaron valores experimentales en un rango de 2.808 a 17.632%. El análisis de regresión mostró que el ϵ fue afectada por la variable de proceso A (tipo de almidón) ($p \leq 0.0579$). La interpretación de la variable se realizó a través del gráfico representativo de un solo factor ($p \leq 0.10$) (Figura 1b) utilizando el paquete estadístico Desig Expert v6.010 (Stat-Ease, 2003). De manera general se observó que a concentraciones de fibra intermedias (5%) se favorecieron los valores de ϵ principalmente por la inclusión de la fibra (anclaje) dentro de la matriz de almidón debido a una retención de agua en menor proporción de la estructura presente en el almidón nativo de maíz ceroso (Tester y Morrison, 1990).

Modulo de Elasticidad (E). Los valores experimentales en láminas obtenidas por el proceso de extrusión variaron de 8.035 a 25.787 MPa. El Análisis de regresión mostró que el E fue afectado por la interacción concentración de humedad fibra H-F ($p \leq 0.0970$). El modelo de predicción para el E utilizó las variables codificadas:

$$Y_E = +7.08 + 0.70 * X_1 + 0.20 * X_2 - 2.83 * X_{3(1)} + 0.72 * X_{3(2)} + 7.44 * X_1 * X_1 + 2.96 * X_1 * X_2 - 0.76 * X_1 * X_{3(1)} - 0.69 * X_1 * X_{3(2)} - 2.43 * X_2 * X_{3(1)} + 0.67 * X_2 * X_{3(2)}$$

De acuerdo al ajuste del modelo no se dispone de las variables en términos reales. El modelo de predicción utilizado explicó el 79.98% de la variación total ($p \leq 0.1$) para los valores de E. Para láminas a base de almidón acetilado y nativo se presentó una misma tendencia en donde los valores máximos de E (19.074 y 23.69 MPa) se observaron a Humedades de 22.75% y concentraciones de fibra del 10% (Figura 1c), indicándonos que este comportamiento pudo ser debido a un incremento proporcional del movimiento

molecular de las cadenas de almidón, presentándose una mayor interacción entre las cadenas poliméricas en almidones cerosos (Bello-Perez *et al.*, 2002).

Conclusiones. 1) La respuesta σ_{\max} fue afectada por la interacción concentración de humedad-fibra H-F ($p \leq 0.10$) en donde a bajas concentraciones de humedad y fibra (18.25% y 0-2.5%), se incrementó σ_{\max} . 2) ϵ fue afectado por la modificación de almidón A ($p \leq 0.0579$) en donde se observó que ϵ decreció a medida que se incrementó el contenido de fibra F (0-10%), siendo favorecidos al incrementarse la humedad H (18.25%-22.75%). 3) La variable respuesta E fue afectada por la concentración humedad fibra H-F ($p \leq 0.0970$), en donde a medida que se incrementaron los contenidos de fibra F y humedad H fueron mayores los valores de E. 4) A través de la superposición de los gráficos de contorno fue posible la selección del tratamiento W15. 5) A través del comportamiento de la tendencia de los gráficos obtenidos, fue posible el análisis del comportamiento de las láminas obtenidas por extrusión termoplástica.

REFERENCIAS

- American Society for Testing and Materials (ASTM) 2001. Standard Test Methods for Tensile Properties of Thin Plastic Sheeting. ASTM D882-00. In: ASTM Standards on Disc. Philadelphia, ASTM v.08.01. 9p., 2001.
- Bello-Pérez A., Contreras-Ramos M., Romero-Manilla R., Solorza-Feria J y Jiménez-Aparicio A. 2002. Chemical and functional properties of modified starches from banana (*Musa paradisiaca*) Var. Macho. *Agrociencia* 36:169-180.
- Cho, S., Park, J., and Rhee, C. 2002. Properties of laminated films from whey powder and sodium caseinate mixtures and zein layers. *Labbindmittel-Wissenschaft und-Technologie* 35(2):135-139.
- Galicia-García T., Martínez-Bustos F., Jiménez-Arévalo O., and Aguilar-Palazuelos E. 2007. Mechanical Properties of Extruded Biodegradable films of native starch from sugar cane fibre. *Food Engineering Research Developments*. Chapter 9. ISBN:978-1-60021-906-1
- Miles, M., Morris, J., Orford, D., and Ring, D. 1985. The roles of amylose and amylopectin in the gelation and retrogradation of starch. *Carbohydrates Research* . 135: 271-281.
- Landerito, AN, and Wang Ya-Jane. 2005. Preparation and properties of starch phosphates using waxy, common, and high-amylose corn starches. II. Reactive extrusion method. *Cereal Chem*, 82(3):271-276.
- López, Olívia V., García, Maria A., and Zaritzky, Noemi E. 2008. Film forming capacity of chemically modified corn starches. *Carbohydrate Polymers* 73 (1):573-581
- Park, J., Whiteside, S., and Cho Y. 2008. Mechanical and water vapor barrier properties of extruded and heat-pressed gelatin films. *LWT* 41: 692-700.
- San Martín-Martínez., Aguilar-Mendez, M.A., Espinosa Solares, T., Pless, R.C., and Quintana, Z. D. 2004. Starch phosphates produced by extrusion: physical properties and influence on yoguth stability. *Starch/Starke*. 42(1):15-17.
- Stat-Ease 2001. Design-expert version 6.0.5. Stat-Ease, Inc. East 2021 Hennepin Ave., Suite 1991. Minneapolis, MN 55413.
- Tester, R., and Morrison, W. 1990. Swelling and gelatinization of cereal starch. I. Effects of amylopectin, amylase, and lipids. *Cereal Chem*. 67:551-557.

Uso del Modelo de Mezclas para Determinar la Asociación Óptima de Gramíneas-Leguminosa

Rigoberto Castro Rivera¹, Alfonso Hernández Garay², Humberto Vaquera Huerta³, Jorge Pérez Pérez⁴, José de la Paz Hernández Girón⁵, Adrian Raymundo Quero Carrillo⁶, Omar Ramírez Reynoso⁷, Javier Francisco Enríquez Quiroz⁸

¹Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional. Instituto Politécnico Nacional. Unidad Oaxaca. México. rigocastro4@hotmail.com.

²Colegio de Postgraduados en Ciencias Agrícolas-Campus Montecillo. Texcoco, México.

³Colegio de Postgraduados en Ciencias Agrícolas-Campus Montecillo. Texcoco, México.

⁴Colegio de Postgraduados en Ciencias Agrícolas-Campus Montecillo. Texcoco, México.

⁵Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional. Instituto Politécnico Nacional. Unidad Oaxaca. México.

⁶Colegio de Postgraduados en Ciencias Agrícolas-Campus Montecillo. Texcoco, México

⁷Colegio de Postgraduados en Ciencias Agrícolas-Campus Montecillo. Texcoco, México

⁸Campo Experimental Cotaxtla. Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias. Isla, Veracruz, México.

Resumen

El objetivo del estudio fue determinar la mejor asociación de dos gramíneas con una leguminosa sembradas en diferentes proporciones sobre el rendimiento de forraje. Se evaluaron 5 tratamientos que consistieron en la asociación de: 40:30:30% trébol blanco-ovillo-ballico perenne; 40:60% trébol blanco-ovillo; 40:60% trébol blanco-ballico perenne; 40:40:20% trébol blanco-ovillo-ballico perenne y 40:20:40% trébol blanco-ovillo-ballico perenne, los cuales se distribuyeron en 20 unidades experimentales en un diseño en bloques completos al azar con cuatro repeticiones. Las parcelas fueron defoliados por ovinos cada 4 semanas en primavera-verano y 6 semanas en otoño-invierno. Una vez obtenidos los rendimientos anuales, se procedió a analizarlos como un experimento en mezclas (Minitab, 14) y así, obtener la proporción óptima, de cada especie asociada que maximice el rendimiento de forraje. Los resultados muestran que la combinación 40:20:40% de trébol blanco-ovillo-ballico perenne presentó la mayor producción de forraje superando con 52% a la proporción 40:60% de trébol blanco-ovillo que fue la que obtuvo los valores más bajos ($P < 0.05$). Independientemente de la asociación, durante la época de primavera-verano se presentó la mayor producción de forraje, representando 65% del total, y el menor rendimiento en otoño e invierno con un 35% del total. Las mejores asociaciones fueron 40:20:40% y 40:30:30% de trébol blanco-ovillo-ballico perenne. Del análisis de mezclas se obtuvo que la proporción ideal que maximiza el rendimiento de forraje es 40% trébol blanco, 22.8% de pasto ovilla y 36.6% de ballico perenne. Dicho valor se obtuvo con la siguiente ecuación de regresión: Rendimiento = 11906 (% pasto ovilla) + 14913 (% ballico perenne) + 12770 ((% pasto ovilla) * (% ballico perenne)).

Palabras clave: Praderas asociadas, Análisis de mezclas, trébol blanco, ballico perenne y ovilla, optimización, superficie de respuesta.

Introducción

El aprovechamiento adecuado de praderas puras o mixtas requiere del conocimiento sobre el manejo agronómico, distribución estacional del rendimiento y la respuesta a la defoliación. Sin embargo, el rendimiento de las plantas forrajeras está sujeto a un patrón de crecimiento estacional que es influenciado por la temperatura, humedad, radiación solar y nutrientes, que determinan la cantidad de biomasa por ciclo de producción, rendimiento estacional y anual (Matthew *et al.*, 2001; Lemaire, 2001). La composición botánica y morfológica son variables que proporcionan una idea del comportamiento de las especies en una asociación, cual domina la cubierta vegetal en los diferentes estratos (Karsten *et al.*, 2002). El uso de leguminosas solas o asociadas con gramíneas permite mejorar el rendimiento y calidad nutricional del forraje, ganancias de peso, productividad en el ganado y la fertilidad del suelo por el aporte de nitrógeno atmosférico (Karsten *et al.*, 2002). Similarmente la persistencia bajo pastoreo ha sido el factor clave para el uso de leguminosas (Quero *et al.*, 2007). Por lo anterior, el objetivo del presente estudio fue evaluar el rendimiento anual de forraje y con el modelo de análisis de mezclas determinar la combinación óptima que maximice el rendimiento.

Materiales y Métodos

El estudio se realizó en una pradera mixta de trébol blanco (*Trifolium repens*), pasto ovillo (*Dactylis glomerata*) y ballico perenne (*Lolium perenne*), ubicada en el campo experimental del Colegio de Posgraduados, en Montecillo, Texcoco, Estado de México. El clima es templado subhúmedo con lluvias en verano, precipitación y temperatura media anual de 645 mm y 15 °C, respectivamente (García, 1988). La proporción de leguminosa-gramíneas fue de 40-60% (White y Hodgson, 1999). Los tratamientos consistieron en las combinaciones: 40:30:30% de trébol blanco - (TB) -ovillo (O) – ballico perenne (BP); 40:60% TB-O; 40:60% TB-BP; 40:20:40% TB-O-BP) y 40:40:20% TB-O-BP, y se distribuyeron en un diseño en bloques al azar con cuatro repeticiones. La frecuencia de defoliación fue cada cuatro semanas en primavera -verano y seis semanas en otoño-invierno, de acuerdo con Velasco *et al.* (2001 y 2005), para praderas de ovillo y ballico perenne en monocultivo.

VARIABLES ESTUDIADAS

Para obtener el rendimiento de forraje, se establecieron dos cuadros fijos de 0.25 m², donde se cortó el forraje presente antes del pastoreo a 5 cm de altura. El material cosechado se lavó y secó en una estufa de aire forzado por 48 horas a 55 °C, hasta peso constante y se pesó. Los datos obtenidos fueron analizados utilizando LSMEANS a un nivel de significancia del 5%, mediante el procedimiento estadístico PROC MIXED del software SAS (SAS, 2002). Además, se realizó un análisis estadístico de mezclas, para determinar la combinación ideal que maximice el rendimiento de forraje, utilizando el software estadístico Minitab ver.14 (MINITAB, 2006).

Resultados y Discusiones

Rendimiento de Forraje

A excepción de la asociación 40:60% TB-O, que registro el menor rendimiento anual, no se presentaron diferencias estadísticas entre tratamientos (Cuadro 1). El rendimiento anual mostró el siguiente orden descendente: 40:20:40% TB-O-BP > 40:30:30% TB-O-BP > 40:60% TB-BP > 40:40:20% TB-O-BP > 40:60% TB-O, con una diferencia del 52% entre

la asociación con el mayor y menor rendimiento de forraje ($P < 0.05$). Las asociaciones 40:20:40% y 40:30:30% de TB-O-BP registraron los valores más altos de producción de forraje, con un aporte de 47 y 49% para el trébol blanco, seguida del ballico perenne con el 23 y 27% y el ovido con 9 y 12%, respectivamente. Los resultados muestran que el 65% del rendimiento anual se concentró en las épocas de primavera y verano, y el 35% en otoño e invierno, siendo la estación de otoño la que presentó el menor rendimiento con 12% ($P < 0.05$). El mayor rendimiento durante las estaciones de primavera – verano lo registró la combinación 40:20:40% de trébol blanco-ovido-ballico perenne, en tanto que la asociación 40:30:30% de trébol blanco-ovido-ballico perenne lo presentó en las épocas de otoño e invierno.

Independientemente de la asociación, el comportamiento estacional del rendimiento de forraje fue de 34, 29, 12 y 22 % para verano, primavera, otoño e invierno, respectivamente, y se atribuyen a que, en primavera - verano se registraron las condiciones ambientales adecuadas, particularmente temperaturas, que permitieron a las especies manifestar su máximo potencial. Al respecto, se ha señalado la temperatura ejerce influencia directa sobre la tasa de aparición y expansión de las hojas y aunado a un nivel adecuado de humedad, las praderas alcanzan rápidamente un dosel cerrado en primavera y verano, coincidiendo con los resultados obtenidos por Velazco *et al.* (2001; 2005), para el pasto ovido y ballico perenne, y Hernández *et al.* (1997) en la asociación trébol blanco- ballico perenne.

Análisis de mezclas

El análisis estadístico de los datos experimentales se realizó usando el modelo para el diseño de mezclas:

$$Y = b_1X_1 + b_2X_2 + b_3X_1X_2 + e$$

Sujeto a la restricción $X_1+X_2=1$ donde X_1 representa la proporción de pasto ovido y X_2 representa la proporción de ballico perenne, Y es la respuesta observada y e es el término de error aleatorio. Usando el software Minitab 14, se estimaron los parámetros del modelo resumiendo los resultados en la siguiente ecuación de regresión:

$$\text{Rendimiento} = 11906X_1 + 14913X_2 + 12770X_1X_2$$

Usando la ecuación anterior se obtuvo la siguiente combinación, que maximiza el rendimiento de materia seca: 40% trébol blanco (valor fijo en la mezcla), 22.8% de pasto ovido y 36.6% de ballico perenne (Cuadro 2).

Conclusiones

El mayor rendimiento anual y durante las estaciones de primavera – verano lo registró la combinación 40:20:40% de trébol blanco-ovido-ballico perenne. La asociación 40:30:30% de trébol blanco-ovido-ballico perenne presentó el más alto rendimiento de forraje en las épocas de otoño e invierno. La combinación que maximiza el rendimiento de forraje derivado del análisis de mezclas fue: 40% trébol blanco, 22.8% de pasto ovido y 36.6% de ballico perenne

REFERENCIAS

- García E. 1988. Modificaciones al sistema de clasificación climática de Köppen. 4ª ed. Universidad Nacional Autónoma de México.
- Hernández Garay A., Hodgson J. and Matthew C. 1997. Effect of spring grazing management on perennial ryegrass and white clover pastures. 1. Tissue turnover and herbage accumulation. *New Zealand J. Agric. Res.* 40:25-35.
- Lemaire G. 2001. Ecophysiology of grasslands: Dynamics aspects of forage plant populations in grazed swards. Proc. XIX International Grassland Congress. Brazilian Society of Animal Husbandry. Sociedade Brasileira de Zootecnia. Sao Pedro, Sao Paulo. Brasil. 29-37.
- Matthew, C. G., E. N. Val Loo, E. R. Tom, L. A. Dawson, and D. A. Care. 2001. Understanding shoot and root development. Proc. XIX International Grassland Congress. Brazilian Society of Animal Husbandry. Sociedade Brasileira de Zootecnia. Sao Paulo, Brasil. 19-27.
- Minitab 2006. Meet Minitab, manual for the basic practice of statistics. Ed. Freeman W. H. USA.
- Quero, C. A. R., J. F. Enríquez Q. y L. Miranda Jiménez. 2007. Evaluación de especies forrajeras en América tropical, avances o *status quo*. Recibido el 8 de Junio. *Interciencia-Venezuela*. 32 (8): 566-571
- Karsten H. D., Carlsson M. 2002. Describing the botanical composition of a mixed species northeastern U.S. Pasture rotationally grazed by cattle. *Crop Sci.* 42:882-889.
- SAS. 2002. SAS User's Guide: Statistics (version 9.0 ed.). Cary NC, USA: SAS Inst. Inc.
- Velasco Z.M.E., Hernández-Garay A., González Hernández V.A., Pérez P. J., Vaquera H. H., Galvis S. A. 2001. Curva de crecimiento y acumulación estacional del pasto ovillo (*Dactylis glomerata* L.). *Téc. Pec. Méx.* 39 (1):1-14.
- Velasco Zebadúa, M. E., Hernández-Garay A., González Hernández V. A. 2005. Rendimiento y valor nutritivo de Ballico perenne (*Lolium perenne* L.) en respuesta a la frecuencia de corte. *Téc. Pec. Méx.*; 43(2):274:258.
- White J, Hodgson J. 1999. Pasture establishment. *In: Pasture and crop science*. Oxford University Press. New Zealand. Pag. 104-106.

Cuadro 1. Rendimiento de forraje (Kg MS ha⁻¹) estacional y anual de cinco asociaciones de gramíneas y leguminosas.

| Estación del año | Asociaciones | | | | | EEM | Sig. | Prom. |
|-------------------|----------------------|-----------------|----------------|----------------------|----------------------|-----|------|--------|
| | 40:30:30% TB-O-BP | 40:60% TB-BP | 40:60% TB-O | 40:20:40% TB-O-BP | 40:40:20% TB-O-BP | | | |
| Primavera | 4514 AB b | 4813 AB a | 3330 B ab | 5137 A b | 4434 AB b | 493 | * | 4446 b |
| Verano | 5803 A a | 5464 A a | 4239 B a | 6096 A a | 5329 AB a | 370 | * | 5386 a |
| Otoño | 2319 A c | 1863 AB c | 1336 B c | 2088 A d | 1725 AB d | 237 | * | 1866 d |
| Invierno | 4149 A b | 3347 AB b | 2452 C bc | 3954 AB c | 3498 AB c | 404 | * | 3480 c |
| Promedio | 4196 AB | 3872 AB | 2839 C | 4318 A | 3746.4 B | 195 | ** | |
| EEM | 271 | 296 | 34 | 242 | 206 | | | 174 |
| Sig. | ** | ** | ** | ** | ** | | | ** |
| Rendimiento anual | 16785 A | 15488 A | 12136 B | 17274 A | 14986 A | 953 | * | |

Diferente literal mayúscula, en cada hilera, indican diferencias significativas (P<0.05), de las asociaciones en cada estación del año

Diferente literal minúscula, en cada columna, indican diferencias significativas (P<0.05), de las asociaciones en las diferentes estaciones del año

TB= Trébol blanco, O= Ovillo, BP= Ballico perenne.

EEM= Error estándar de la media.

Sig.= Significancia.

Prom.= Promedio.

Cuadro 2. Análisis de mezclas y obtención de la ecuación de regresión que maximiza el rendimiento de forraje.

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|----------------|----|-----------|-----------|----------|-------|-------|
| Regression | 2 | 124412918 | 124412918 | 62206459 | 12.85 | 0.000 |
| Linear | 1 | 40344568 | 40344568 | 40344568 | 8.33 | 0.006 |
| Quadratic | 1 | 84068350 | 84068350 | 84068350 | 17.36 | 0.000 |
| Residual Error | 37 | 179137759 | 179137759 | 4841561 | | |
| Lack-of-Fit | 2 | 49332106 | 49332106 | 24666053 | 6.65 | 0.004 |
| Pure Error | 35 | 129805653 | 129805653 | 3708733 | | |
| Total | 39 | 303550676 | | | | |

| Term | Coef | SE | Coef | T | P | VIF | ovillo |
|----------------|-------|--------|------|-------|-------|-----|--------|
| 11906 | 756.5 | * | * | 1.709 | | | |
| ballico | 14913 | 756.5 | * | * | 1.709 | | |
| ovillo*ballico | 12770 | 3064.5 | 4.17 | 0.000 | 2.487 | | |

Una actividad conducente a lograr el aprendizaje significativo del concepto de error tipo I

Félix Almendra-Arao¹, David Sotres-Ramos², Cecilia Ramírez Figueroa³

¹UPIITA del Instituto Politécnico Nacional. falmendra@ipn.mx

²Programa de Estadística. Colegio de Postgraduados. sotres.davida@kendle.com

³Programa de Estadística. Colegio de Postgraduados. figueroa@gmail.com

Resumen.

La adecuada construcción por parte del estudiante del concepto de error tipo I de una prueba de hipótesis es esencial para la comprensión de la teoría de prueba de hipótesis, ya que éste concepto es una pieza fundamental en dicha teoría. No obstante, es usual que el estudiante al tratar de comprender este concepto atravesase por dificultades en el proceso de construcción de su significado. Por lo anterior, en este trabajo describimos una actividad para auxiliar al estudiante en la construcción del concepto de error tipo I (α) de una prueba de hipótesis estadística, en cursos introductorios de estadística. La actividad se realiza en el contexto de la comparación de dos proporciones independientes. La interpretación resultante de esta actividad por parte del estudiante, para α , después del proceso de construcción del concepto es: “la proporción de experimentos que por azar nos permiten declarar diferencia entre tratamientos cuando en realidad no existe esta diferencia”. La actividad que presentamos se fundamenta desde el punto de vista teórico en la teoría APOS y el ciclo de enseñanza ACE, dentro del marco de la teoría APOS.

I. Introducción

La estadística es una de las asignaturas más enseñadas a nivel universitario ya que se encuentra presente en los mapas curriculares de carreras en áreas del conocimiento tan diversas como humanidades, ciencias biológicas, ciencias sociales y ciencias exactas. Tan sólo debido a su amplia enseñanza, se justifica la realización de investigación acerca de su enseñanza; la investigación acerca del pensamiento estocástico tiene sus orígenes en los trabajos de Piaget en los años 50's, Piaget et Inhelder (1951), del siglo pasado. En recientes trabajos de investigación, (ver por ejemplo, Batanero (2000), Batanero and Díaz (2006), Batanero (2007), Cumming, Williams and Fidler (2004), Díaz y Batanero (2006), Díaz y de la Fuente (2004), Falk (1986), Hager (2000), Haller and Krauss (2002), Menon (1993), Pascual, García y Frías (2000)), se plantea claramente la necesidad de realizar investigación acerca de educación estadística debido a que existe una amplia variedad de problemáticas relacionada con la enseñanza de la estadística. Específicamente, en inferencia estadística, es usual que el estudiante desarrolle concepciones equivocadas relativas a pruebas estadísticas, particularmente en lo que se refiere al nivel de significancia o error tipo I (α), el cual se define como la probabilidad de rechazar la hipótesis nula cuando esta es verdadera.

“En años recientes, una comunidad de investigadores ha estado usando y refinando un particular marco de trabajo para investigación y desarrollo curricular en educación matemática a nivel universitario”, ver Asiala *et al.* (1996), dando con ello origen a la teoría

APOS, ver Dubinsky and McDonald (2001). “La teoría APOS parte de la premisa de que el conocimiento matemático consiste en una tendencia individual a tratar con situaciones de problemas matemáticos percibidas por medio de construcciones mentales: *acciones, procesos, objetos y esquemas*, para hacer sentido de las situaciones y resolver los problemas, en referencia a esas construcciones mentales nosotros le llamamos teoría APOS. Esas ideas provienen de nuestros intentos por extender al nivel del aprendizaje de matemáticas de preparatoria, el trabajo de J. Piaget sobre abstracción reflexiva en el aprendizaje del niño” (Dubinsky and McDonald (2001).

Desde el punto de vista de la teoría APOS un concepto puede descomponerse genéticamente para establecer los niveles de Acción, Proceso, Objeto y Esquema. Un acercamiento pedagógico particular, el cual es usado comúnmente dentro del marco de la teoría APOS es el ciclo de enseñanza ACE (Actividades, Clases y Ejercicios). (ver Asiala et al. (1996)), el cual no necesariamente es una consecuencia del marco teórico mencionado, pero es un diseño global el cual soporta el análisis teórico en dicho marco.

Desde esta perspectiva, se diseña una actividad para ser utilizada en el aula, la cual ha mostrado ser un instrumento efectivo para la comprensión de un concepto fundamental en estadística, el error tipo I. En este trabajo se presenta una actividad que puede utilizarse para ayudar al estudiante a comprender el concepto error tipo I (α) de una prueba de hipótesis estadística. Esta actividad puede ser utilizada en cursos introductorios de estadística para estudiantes en las de diversas áreas de conocimiento como ciencias biológicas, ciencias sociales, etc.

La actividad está diseñada para ser realizada en el contexto de la comparación de dos proporciones independientes utilizando una muestra aleatoria de cada una de dos poblaciones Bernoulli.

En esta actividad se emplean dos urnas de canicas que los propios alumnos utilizan para generar un gran número de repeticiones de este experimento (parejas de muestras aleatorias de poblaciones Bernoulli), de preferencia generar 100 experimentos. Cada uno de estos experimentos es analizado por los mismos estudiantes utilizando la prueba ji-cuadrada, posteriormente con los resultados obtenidos de los 100 experimentos se obtiene una estimación y también una interpretación significativa del error tipo I (α). La interpretación resultante de esta actividad por parte de los estudiantes es que el error tipo I significa: “la proporción de experimentos que por azar conducen a una conclusión errónea, declarar diferencia significativa entre tratamientos cuando en realidad no existe esta diferencia”.

La experiencia indica que esta actividad es muy efectiva para que el estudiante comprenda de manera significativa, es decir, de tal forma que el concepto adquiera un significado para él e integre a su estructura cognitiva de manera sustancial el concepto de error tipo I.

En el presente trabajo se presenta una forma en la cual se puede contribuir a un mejor entendimiento del concepto de error tipo I.

II. Modelo Estadístico Empleado en la Actividad

El modelo empleado es el de dos poblaciones Bernoulli con parámetros P_1 y P_2 respectivamente, que representan la probabilidad de respuesta positiva del tratamiento estándar y del tratamiento nuevo, respectivamente. El problema consiste en contrastar las siguientes hipótesis

$$H_0 : P_1 = P_2 \quad \text{vs} \quad H_a : P_1 \neq P_2 \quad (1)$$

Utilizando dos muestras aleatorias de estas dos poblaciones Bernoulli, es decir X_1, X_2, \dots, X_m es una muestra aleatoria de la población Bernoulli con parámetro P_1 y Y_1, Y_2, \dots, Y_n es una muestra aleatoria de la población Bernoulli con parámetro P_2 . La prueba que se utiliza frecuentemente para contrastar las hipótesis en (1) es la prueba asintótica de ji-cuadrada, ver por ejemplo Glantz (1992). Esta prueba para un error tipo I (α), rechaza la hipótesis nula en favor de la alternativa cuando

$$(\chi^2)_c > \chi^2(1, \alpha) \quad (2)$$

en dónde,

$$(\chi^2)_c = \left\{ \left| p_1 - p_2 \right| - (1/m + 1/n)/2 \right\}^2 / p(1-p)(1/m + 1/n), \quad (3)$$

con $p_1 = \sum X_i / m$, $p_2 = \sum Y_i / n$, y $p = (m p_1 + n p_2) / (m + n)$, y $\chi^2(1, \alpha)$ es el percentil de nivel $1-\alpha$ de la distribución ji-cuadrada con un grado de libertad.

III. Etapas de la Actividad

La actividad propuesta se desglosa en tres etapas, mismas que se describen a continuación.

ETAPA#1 Generación de los datos. La etapa 1 de la actividad consiste en que los alumnos generen 100 experimentos, o sea 100 parejas de muestras aleatorias de dos poblaciones Bernoulli con parámetros P_1 y P_2 respectivamente, utilizando dos urnas de canicas con 100 canicas cada una. Por supuesto, las extracciones de las urnas de canicas deben realizarse con reemplazo. Debemos enfatizar que lo que se desea es estimar el valor de α , o sea deseamos estimar a la $\text{Prob} \{ \text{Rechazar } H_0 : P_1 = P_2 \mid H_0 \text{ es verdadera} \}$, por lo que, tomando como ilustración $P_1 = 0.85$, las 2 urnas se llenan de manera idéntica con 85 canicas blancas (éxito) y con 15 canicas negras (fracaso). De esta manera tenemos que para las 2 poblaciones se cumple que $P_1 = P_2 = 0.85$, y por tanto la hipótesis nula es la verdadera. Sin embargo, en la actividad procederemos como si no supiéramos esto, es decir para cada experimento generado se realiza la prueba de ji-cuadrada con los datos obtenidos de las urnas de canicas. Al final de esta etapa obtenemos para cada uno de los 100 experimentos: el número de éxitos observados en cada una de las dos muestras extraídas de las urnas. Para generar las muestras conviene dividir el trabajo entre todos los estudiantes, por ejemplo con 20 estudiantes, entonces cada estudiante generaría 5 experimentos. Los tamaños de muestra que hemos observado en la práctica que dan buenos resultados en la clase son $m=n=50$.

ETAPA#2 Análisis de los 100 experimentos. Para cada uno de los 100 experimentos generados los estudiantes realizan la prueba de ji-cuadrada en (2). Siguiendo con el ejemplo de 20 estudiantes, cada estudiante analizaría los 5

experimentos que el mismo generó. Para efectuar la prueba de ji-cuadrada conviene realizarla con un paquete de cómputo sencillo para acortar el tiempo de la actividad.

ETAPA#3 Estimación de α . La etapa final, con los resultados de los 100 experimentos analizados, consiste en elaborar una tabla que concentra los resultados de estos experimentos.

Tabla 1. Concentrado de resultados de los 100 experimentos

| | |
|---|-------|
| Número de experimentos que rechazaron la hipótesis nula | k |
| Porcentaje de experimentos que rechazaron la hipótesis nula | k/100 |

Con los resultados de la Tabla 1 se obtiene como estimador de α a $k/100$, es decir el porcentaje de experimentos que por azar rechazaron la hipótesis nula cuando ésta en realidad es cierta. Usualmente, al finalizar la actividad, el valor que resulta para $k/100$ es igual a 0.03, 0.04, 0.05, 0.06, ó 0.07. Se explica a los estudiantes que si n es el número de experimentos y hacemos crecer el número de experimentos entonces $k/n \rightarrow 0.05$ cuando $n \rightarrow +\infty$. Una ventaja didáctica adicional de la actividad es que al elaborar una lista de los 100 valores observados de la estadística de prueba de la ji-cuadrada en (3) es posible elaborar, con estos valores observados, un histograma sencillo que mostrará la forma de la gráfica de la función de densidad de la estadística ji-cuadrada con 1 grado de libertad, y observar en la gráfica cuales fueron los experimentos que rebasaron la ji-cuadrada de tablas y nos llevaron a tomar una decisión errónea.

IV. Conclusiones

La actividad descrita tiene la ventaja de que se está utilizando un método constructivista, es decir, son los propios estudiantes quienes generan, concentran y analizan los datos de los 100 experimentos, y con base en los resultados de estos experimentos obtienen como estimación de α al porcentaje de experimentos que por azar rechazaron la hipótesis nula cuando en realidad es cierta, mediante lo cual construyen su concepto de error de tipo I. La experiencia indica que esta actividad es altamente efectiva para que el estudiante comprenda de tal forma que el concepto adquiriera un significado para él e integre a su estructura cognitiva de manera sustancial el concepto así como la interpretación de error tipo I. Además, otra ventaja didáctica de la actividad es que produce un histograma sencillo que mostrará la forma de la gráfica de la función de densidad de la estadística ji-cuadrada con 1 grado de libertad.

REFERENCIAS

- Asiala, M., Brown, A., De Vries, D., Dubinsky, E., Mathews, D. and Thomas, K. (1996). A Framework for Research and Curriculum Development in Undergraduate Mathematics Education. *Research in Collegiate Mathematics Education II, Issues in Mathematics Education (CBMS)*, American Mathematical Society.
- Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, **2**, (1-2), 75 – 98.
- Batanero, C. and Díaz, C. (2006). Methodological and Didactical Controversies around Statistical Inference. *Actes du 36ièmes Journées de la Société Française de Statistique*. CD ROM. Paris: Société Française de Statistique.
- Batanero, C. (2007). Emerging research in statistics education. *International Electronic Journal of Mathematics Education* **2**, (3), 107-109.
- Cumming, G.; Williams, J. and Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics* **3**, 299-311.
- Díaz, C. y Batanero, C. (2006) ¿Cómo puede el método bayesiano contribuir a la investigación en psicología y educación? *Paradigma*. **27**, (2), 35-54.
- Díaz, C. y de la Fuente, I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento*, Volumen especial, 161-167.
- Dubinsky, E. and McDonald, M. (2001). APOS: A Constructivist Theory of Learning in Undergraduate Mathematics Education Research. In D. Holton et. (Eds.), *The teaching and Learning of Mathematics at University Level: An ICMI Study*, Kluwer Academic Publishers, 273-280.
- Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning* **9**, 83 – 96.
- Glantz, S. (1992). *Primer of Biostatistics*, 3rd Edition, Mc Graw - Hill Inc.
- Hager, W. (2000). About some misconceptions and the discontent with statistical tests in psychology. *Methods on Psychological Research*, **5**, (1). 1-31. En línea: <http://www.dgps.de/fachgruppen/methoden/mpr-online/issue9/art1/hager.pdf>
- Haller, H. and Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research*, **7**, (1), 1-20. En línea: <http://www.mpronline.de/issue16/art1/haller.pdf>
- Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal*, **5**, (1), 4 – 18.
- Pascual, J., García, J.F. y Frías, M.D. (2000). Significación estadística, importancia del efecto y replicabilidad de los datos *Psicothema*, **12**, Suplemento 2, 408-412
- Piaget, J. et Inhelder, B. (1951). *La genèse de l'idée de hasard chez l'enfant*. Presses Universitaires de France, Paris.

Una experiencia Docente para el Aprendizaje de Modelos Probabilísticos Discretos Basadas en la Redacción de Enunciados de Problemas

M. J. García-Ligero¹, P. Román-Román²

¹Dpto. de Estadística e I.O. Universidad de Granada. mjgarcia@ugr.es

²Dpto. de Estadística e I.O. Universidad de Granada. proman@ugr.es

Resumen

La forma clásica de ejercitar al alumno en el conocimiento y la aplicación práctica de los modelos probabilísticos discretos más usuales es el planteamiento de problemas para su resolución, una vez estudiada de forma teórica las características más importantes de cada uno de ellos, junto a las similitudes y diferencias con el resto.

En este trabajo se presenta una experiencia docente en la que hemos dado un giro a esta dinámica de estudio y el alumno, una vez estudiados de forma teórica los modelos, no se va a enfrentar con problemas totalmente estructurados para limitarse simplemente a deducir el modelo adecuado a la vista del cumplimiento de las hipótesis correspondientes. Por el contrario, se le va a solicitar que redacte enunciados de problemas que permitan que el resto de compañeros sean capaces de identificar, sin ningún género de dudas, el modelo probabilístico al que corresponde.

1. Introducción

En los sistemas clásicos de enseñanza, la forma habitual de ejercitar al alumno en el reconocimiento de los distintos modelos probabilísticos discretos en situaciones reales concretas ha sido el planteamiento de problemas para su resolución, una vez estudiada de forma teórica las características más importantes de cada uno de ellos, junto a las similitudes y diferencias con el resto.

Sin embargo, este sistema de enseñanza en la que el alumno se limita a ir comprobando la verificación o no de las hipótesis de los distintos modelos en los enunciados de los problemas propuestos presenta algunas deficiencias. Entre ellas podemos citar el hecho de que el alumno detecta el modelo más por eliminación de otros que por la completa comprensión de las características del modelo involucrado en su caso real. Todo ello nos ha llevado a plantearnos una experiencia docente en la que hemos dado un giro a esta dinámica de estudio y el alumno, una vez estudiados de forma teórica los modelos, no se va a enfrentar con problemas totalmente estructurados para limitarse simplemente a deducir el modelo adecuado a la vista del cumplimiento de las hipótesis correspondientes. Por el contrario, se le va a solicitar que redacte enunciados de problemas de forma que se vea involucrado en la comprensión de la naturaleza del problema de tal forma que todas las características e hipótesis queden especificadas tan claramente que sus compañeros sean capaces de identificar, sin ningún género de dudas, el modelo probabilístico al que corresponde.

2. Experiencia

Esta experiencia se ha llevado a cabo con un grupo de alumnos de la asignatura "Probabilidades y Estadística" de segundo curso de la Licenciatura en Matemáticas de la Universidad de Granada durante el curso académico 2008-2009.

Así, en primer lugar se explicó en clase, de forma teórica, las características más importantes de los principales modelos probabilísticos discretos siguiendo por ejemplo [1]. En concreto, para esta experiencia nos centramos en los siguientes modelos:

- Modelo binomial:
 - Existencia de un experimento aleatorio de tipo Bernoulli (con dos posibles resultados: éxito y fracaso).
 - Número (fijado de antemano (n)) de repeticiones independientes del experimento con probabilidad de éxito (p) constante.
 - Variable de interés: Número de éxitos en las n repeticiones independientes del experimento.
- Modelo binomial negativo:
 - Existencia de un experimento aleatorio de tipo Bernoulli.
 - Repeticiones independientes del experimento con probabilidad de éxito (p) constante hasta conseguir un número (fijado de antemano (k)) de éxitos.
 - Variable de interés: Número de fracasos en las repeticiones independientes del experimento realizadas antes de conseguir el k -ésimo éxito.
- Modelo hipergeométrico

- Existencia de una población de tamaño N en las que hay N_1 individuos con una cierta característica y, por tanto, $N - N_1$ que no la poseen.
- Realización de un muestro sin reemplazamiento o simultáneo de n individuos de la población.
- Variable de interés: Número de individuos que presenta la característica bajo estudio en la muestra de tamaño n .

A continuación se planteó la experiencia didáctica que constó de tres fases fundamentales

1. Redacción de enunciados.

En esta fase, los alumnos (de forma individual o por grupos) se plantearon, en primer lugar, la búsqueda de situaciones prácticas reales en las que aparezcan los tres modelos probabilísticos discretos estudiados. A continuación, asociadas a dichas situaciones, redactaron enunciados de problemas prestando especial interés en plasmar de forma clara todas las cuestiones importantes de cara a identificar las hipótesis requeridas.

Dichos enunciados fueron entregados a las profesoras las cuales, sin modificar nada, elaboraron un documento de trabajo compuesto por todos los enunciados que se entregó a todos los alumnos que participaron en la experiencia.

2. Lectura y discusión pública de los enunciados.

Esta fase se realizó en sesiones de clase donde, a partir de la lectura de cada uno de los enunciados, se mantenía una discusión sobre la verificación o no de las hipótesis de cada uno de los modelos estudiados. Así, se prestó especial interés en comentarios relativos al cumplimiento de hipótesis tales como: existencia de un experimento de Bernoulli que se puede repetir en iguales condiciones; independencia, en su caso, de los experimentos de Bernoulli; existencia de una población de tamaño finito con una propiedad que la dicotomiza; tipo de muestreo realizado en ella; equivalencia de la selección de un individuo de tal población con la realización de un experimento de tipo Bernoulli; pérdida de la independencia en la repetición del experimento consistente en la selección de un individuo en el caso de realizar un muestreo con reemplazamiento en una población finita, etc.

3. Resolución, individual o por grupos, de los problemas propuestos.

Una vez, realizada la fase previa, los alumnos realizaron la fase de resolver los problemas como trabajo complementario al de clase, también de forma individual o por grupos.

3. Un ejemplo de enunciado

Para finalizar, vamos a presentar uno de los enunciados que se propusieron, comentando las principales discusiones a que dió lugar:

“Un niño entra en una tienda de chucherías y quiere comprar gominolas de un recipiente que tiene el 73% de las gominolas de fresa y el resto de limón.

a) Si el niño compra 40 gominolas, ¿cuál es la probabilidad de que haya, al menos, 20 gominolas de limón en la compra?, ¿cuál es el número esperado?

b) Sabiendo que la bolsa tiene 23 gominolas de fresa y 17 de limón, calcular la probabilidad de que la primera gominola de fresa sea la quinta que se come el niño

c) En las condiciones del apartado anterior, si de la bolsa se toman 10 gominolas al azar, calcular la probabilidad de que 7 de ellas sean de limón.”

Las discusiones planteadas fueron dirigidas fundamentalmente a las siguientes cuestiones:

- Existencia de un experimento de Bernoulli (extracción de una gominola del recipiente) con dos posibles resultados (gominola de fresa o gominola de limón).
- Existencia y tamaño de una población (de gominolas) con una propiedad que la dicotomiza (sabor de la gominola).
- Extracción sin reemplazamiento de las gominolas del recipiente.

Así, con respecto al apartado a) supuesto que el muestreo se realizaba sin reemplazamiento y había una población con una propiedad que la dicotomiza se planteó la posibilidad de utilizar un modelo hipergeométrico. Sin embargo, no se disponía del dato del tamaño de la población. Además el tipo de muestreo asociado hacía que en las repeticiones de un experimento de Bernoulli la probabilidad de éxito no se mantuviera constante a menos que se considerara una población infinita en el que las proporciones de gominolas de ambos sabores se mantuvieran siempre constantes (por ejemplo, un rellenado automático del recipiente con una gominola del mismo tipo de la que se extrae). En este último caso, sería posible la utilización de un modelo binomial.

En el apartado b) el planteamiento iba dirigido, en principio, a un modelo binomial o binomial negativo; sin embargo, el tipo de muestreo considerado invalida la consideración de estos modelos en esta situación.

Por último, el apartado c) puede ser resuelto mediante un modelo hipergeométrico al estar fijado el tamaño de la población.

4. Conclusiones

Con esta experiencia se han conseguido una serie de objetivos, con respecto al desarrollo de la docencia, entre los que destacamos:

- Incentivar la participación del alumno en la clase y el trabajo en grupo.
- Promover un aprendizaje crítico de los contenidos explicados en clase.

Por otra parte, algunas ventajas de este nuevo planteamiento, frente al método clásico de enseñanza anteriormente citado son:

- Una mejor asimilación de las hipótesis asociadas a los modelos entre los que destacamos la independencia de los experimentos y la distinción, junto con las consecuencias que acarrea, entre muestreo con y sin reemplazamiento, etc.
- Una mejora en la adquisición de destrezas en la identificación de los modelos.

Referencias

- [1] García-Ligero, M.J., Hermoso, A., Maldonado, J.A., Román, P y Torres, F. (2008) *Curso básico de Probabilidad con CDPYE*. Ed. Copicentro, Granada. ISBN: 978-84-612-2042-7.

Métodos de apoyo/recursos interactivos de aprendizaje de SPSS

Lara-Porras, A.M.¹, Alfonso-Uxó, A.², Jiménez-Lara Pablo²

¹Dpto.de Estadística e I.O. Campus de Fuentenueva. Universidad de Granada. alara@ugr.es

²ETS de Ingeniería Informática y de Telecomunicación. Universidad de Granada.

aleuxo@correo.ugr.es

³Consejería de Innovación, Ciencia y Empresa. Junta de Andalucía. Sevilla.

pablo.jimenez.lara@juntadeandalucia.es

Resumen

Proponemos, un proceso de formación a distancia con un carácter abierto, interactivo y flexible que se desarrolla a través de las nuevas tecnologías de la información y la comunicación, aprovechando sobre todo las bondades que ofrece Internet.

La adopción de un modelo de e-learning supone una apuesta por un modelo pedagógico en el que el alumnado toma una mayor responsabilidad en su educación, contribuyendo a un desarrollo de la eficiencia en el proceso de enseñanza/aprendizaje, y por ende, la mejora cualitativa del modelo educativo. El proceso de interacción en el que se asienta este modelo y el trabajo que presentamos deriva de procesos educativos mas individualizados, más colaborativos.

Por otra parte, teniendo en cuenta la importancia que ha cobrado en nuestra sociedad aspectos como aprendizaje a lo largo de toda la vida, el desarrollo de aprendizajes en red, la autoinformación, etc. El desarrollo de programas interactivos, como el que presentamos, permiten la mejora tanto en la calidad como en la accesibilidad a la educación.

En este trabajo presentamos un sitio Web, <http://www.ugr.es/~bioestad/>, con información tanto general como direccionada a los alumnos a los que va dirigido, material didáctico para las asignaturas de Estadística de la Licenciatura en Biología de la Universidad de Granada y un programa interactivo para el aprendizaje autónomo del alumno del paquete estadístico SPSS. Este soporte lógico esta compuesto de un portal Web <http://www.ugr.es/~bioestad/guiaspss/index.html> donde se pueden localizar diferentes prácticas explicadas y realizadas con el programa SPSS y un emulador del paquete estadístico SPSS.

El portal Web se ha implementado en HTML+JavaScript facilitándose, con este lenguaje de marcas, el acceso a la información ya que este sitio Web es totalmente transparente para la tecnología Web y puede ser visualizado desde cualquier navegador Web (Internet Explorer, Netscape, Mozilla Firefox). Dicho portal Web es la base o el entorno sobre el que se realizan todas las prácticas.

Palabras Claves: Portal Web, Guía interactiva, Autoaprendizaje, TIC, SPSS, Emulador, HTML+JavaScript

1. Introducción

El cambio iniciado en Europa hacia la conformación de un espacio europeo de educación superior implica el establecimiento de nuevos puntos de encuentro entre el profesorado y el alumnado que superen la tangibilidad del aula. En este sentido las Tecnologías de la información y la comunicación (TIC) se conforman como esenciales en el desarrollo de este proceso, en el cual el interés se centra en aumentar el nivel de comprensión por parte del alumnado y reformar, potenciar y reforzar los sistemas pedagógicos tradicionales. La necesidad de adaptación al EEES promueve que el docente planifique una docencia donde las clases magistrales se pueden hacer más dinámicas a través del uso de las TICs. Al mismo tiempo los actuales y continuos progresos tecnológicos han dado lugar a un importante despliegue de estas tecnologías. Este hecho conduce a la modificación de nuestra forma de trabajar, interactuar y al modo de aprendizaje de los estudiantes.

El presente trabajo propone nuevas estrategias y nuevos objetos de estudio relacionados con las creencias, ventajas y desventajas de un modelo educativo basado en la virtualidad que puede favorecer el desarrollo de competencias para el aprendizaje autónomo y una utilización adecuada de las TICs.

2. Descripción

Este proyecto educativo es impulsado por las demandas estadísticas que tiene un titulado en Biología en el campo laboral y profesional. Es necesario que le facilitemos al alumno un “programa” que incluya técnicas estadísticas, resolución de dichas técnicas mediante software apropiado. Es necesario un software que sirva como herramienta para analizar datos con los que el alumno se va a enfrentar en su futuro profesional. A tal fin hemos construido un recurso interactivo de aprendizaje: una guía de autoaprendizaje de SPSS.

El presente trabajo es una continuación y ampliación de un proyecto anterior formado por un portal Web y una programa de autoaprendizaje de SPSS. En este proyecto se amplía, revisa, actualiza la documentación presentada en el portal Web: <http://www.ugr.es/~bioestad/> y se introduce una nueva práctica que facilita y que permite al alumno profundizar en el aprendizaje del SPSS, <http://www.ugr.es/~bioestad/guiaspss/index.html>.

El portal Web, en su página de inicio, presenta información sobre las materias de Bioestadística y Fundamentos de Biología Aplicada I, asignaturas a las que va dirigido el trabajo. Dicho sitio Web consta del material didáctico necesario para el desarrollo de las citadas materias. El objetivo es poner a disposición del alumno tanto en formato tradicional (en papel) como en soporte informático el material apropiado, de esta forma se podrá reducir el tiempo empleado en la exposición teórica de los temas para ampliar la puesta en práctica de resolución de problemas y tratamiento de datos mediante ordenador; introduciendo al alumno en el manejo de paquetes estadísticos. Así mismo, nos muestra el acceso la “Guía interactiva de Autoaprendizaje de SPSS” (Figura 1)

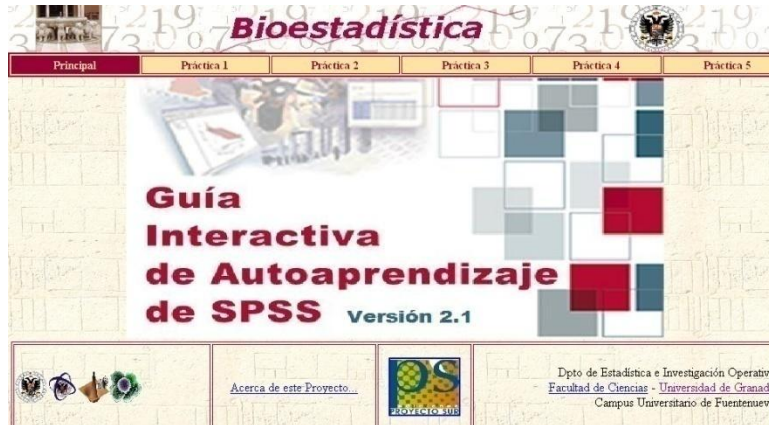


Figura 1: Portal Web (Guía Interactiva de Autoaprendizaje de SPSS)

Este recurso interactivo de autoaprendizaje de SPSS está formado por un Portal Web y un programa emulador del paquete estadístico SPSS:

En el Portal Web (Figura 1) se muestran cinco prácticas basadas en ejercicios con SPSS

- **Práctica 1:** El objetivo es familiarizarse con el entorno del editor de datos de SPSS (Figura 2)
- **Práctica 2:** El objetivo es resumir, ordenar, analizar y representar conjuntos de datos (Figura 3)
- **Práctica 3:** El objetivo es realizar ajustes lineales y parabólicos (Figura 4)
- **Práctica 4:** El objetivo es el estudio de distribuciones (Binomial, Poisson, Normal) (Figura 5)
- **Práctica 5:** El objetivo es el estudio de Intervalos de Confianza (Figura 6).

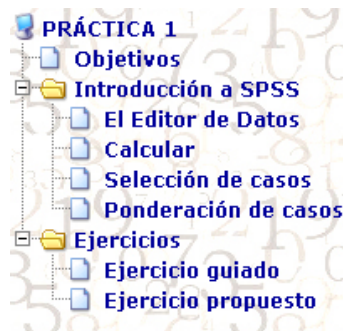


Figura 2: Portal Web (Práctica 1)

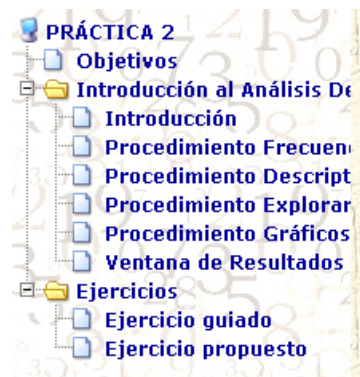


Figura 3: Portal Web (Práctica 2)

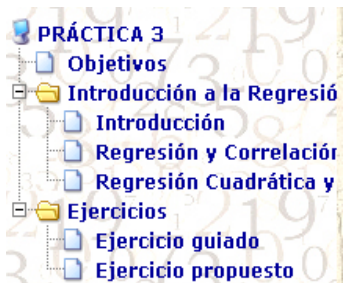


Figura 4: Portal Web (Práctica 3)

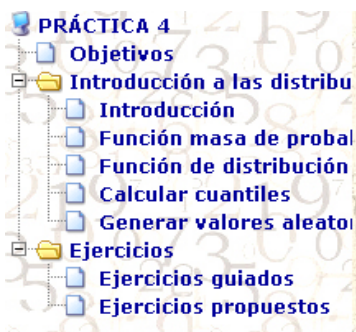


Figura 5: Portal Web (Práctica 4)

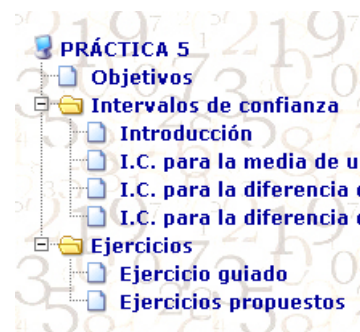


Figura 6: Portal Web (Práctica 5)

Cada práctica contiene una introducción teórica, ejercicios resueltos con SPSS, guiados paso a paso en un entorno similar al utilizado por el propio paquete, empleando para ello un emulador del programa y por último se plantean ejercicios propuestos que el usuario debe realizar directamente con el paquete estadístico.

3. Metodología

Introducimos en las prácticas docentes nuevos métodos de enseñanza constructivistas que contemplan el uso de las TICs como instrumento cognitivo. Estas tecnologías tienen un fuerte impacto en el proceso de enseñanza y se han convertido en complemento y alternativa válida a las clases presenciales. Las TICs constituyen un medio único para potenciar el autoaprendizaje del alumno. En particular, el recurso interactivo de aprendizaje que presentamos ha sido desarrollado utilizando nuevas tecnologías, como el lenguaje de programación multiplataforma JAVA, empleado para desarrollar el programa emulador de SPSS. Asimismo, hemos combinado el *HTML* con *Javascript* para desarrollar el *portal Web* que incluye el material didáctico introductorio y el programa emulador de SPSS.

La aplicación del aprendizaje virtual de este modo, da origen a un tipo de enseñanza basada en la autoinformación y la autoconstrucción del proceso educativo. En este sentido, el trabajo en el aula requiere del trabajo individual del alumno que en todo momento debe estar monitorizado y dirigido. Aparecen nuevas herramientas de trabajo: las plataformas de educación virtual, que abren nuevas posibilidades y ofertas educativas. Éstas son de gran utilidad en la actividad docente puesto que permiten fácilmente el intercambio de material didáctico, pueden emplearse para la realización de tutorías virtuales mediante la pizarra electrónica, así como hacen posible el intercambio de experiencias e inquietudes a través de los foros que disponen. Con nuestros alumnos, la formación virtual está tutelada a través de la plataforma de enseñanza/aprendizaje, *Sistema Web de Apoyo a la docencia* (SWAD <http://swad.ugr.es/>), de la Universidad de Granada. La finalidad de su uso es que los alumnos participen activamente en el desarrollo de la asignatura, para lo cual articulamos diferentes mecanismos de trabajo colaborativo que permitan el proceso de aplicación y utilización de la información que fluye por la red.

Cuando el alumno haya leído toda la documentación e información referente a la práctica que esté realizando (todo esto desde el Portal Web), entra en juego un emulador del

programa SPSS. Con este emulador el alumno empieza a tomar contacto con el programa, ya que su función es guiarlo en la realización de un ejercicio específico de forma interactiva. La siguiente figura muestra el Editor de datos realizado con el programa emulador, donde podemos observar que la interfaz es similar a la del SPSS. Este programa no es una copia del paquete SPSS, ya que no están activas todas las funciones sino sólo las necesarias para la realización de la práctica.

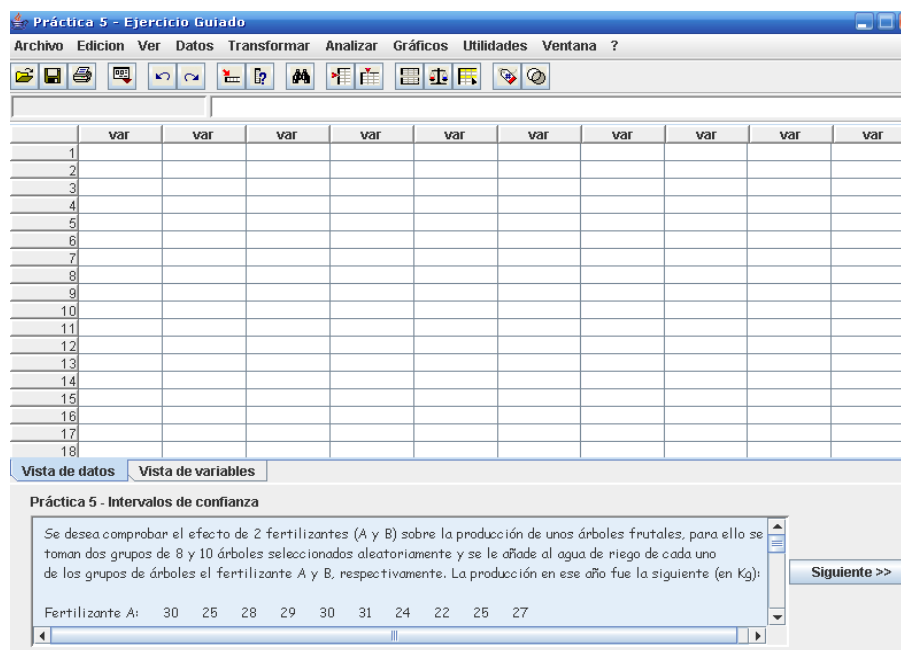


Figura 7. Editor de datos del ejercicio guiado

4. Conclusiones

Creemos necesario e imprescindible que nuestros alumnos utilicen paquetes estadísticos, manejen software estadístico apropiado. El alumno está acostumbrado a una enseñanza “unidireccional.” El profesor es el transmisor de conocimientos y el alumno es el receptor. Es necesario que el alumno se acostumbre a una nueva metodología activa de aprendizaje en la que él, es el principal protagonista del aprendizaje. Por ello debemos dar un impulso decidido a las nuevas tecnologías como las que presentamos en este trabajo.

Con este proyecto, se pretende por una parte continuar el desarrollo de una herramienta informática en el que el principal recurso didáctico a utilizar es un programa interactivo que hace posible el autoaprendizaje del alumno y que les es útil para poder completar su formación en la materia, tanto dentro como fuera del aula y por otra parte que el alumno participe de manera activa en el aprendizaje y la profundización de los contenidos.

En muchas ocasiones, aún con el material teórico y práctico resuelto, el estudio de la Estadística presenta muchos problemas al alumnado cuando se enfrenta de forma personal con ella. Las dudas que se plantean en la realización de ejercicios y en la comprensión de aspectos teóricos, en ocasiones paralizan el avance en el estudio de la asignatura. Por ello, con este programa se pretende que el alumno pueda comprender de forma explícita los

resultados teóricos de la asignatura, sea capaz de ir resolviendo los ejercicios paso a paso comprendiendo la metodología y globalizar su aprendizaje teórico-práctico con los programas computacionales estadísticos que se desarrollan en las clases aplicadas con ordenador en el laboratorio, siendo finalmente autoevaluado cada alumno, por el mismo programa, al final de cada bloque de contenidos.

5. Bibliografía

Buzón García, O. et al. (2005). Estudio sobre la viabilidad de las propuestas metodológicas derivadas de la aplicación del Crédito Europeo por parte del Profesorado de las Universidades Españolas, vinculadas a la utilización de las TICs en la Docencia y en la Investigación. Madrid. *Dirección General de Universidades e Investigación*. 497.

Lara Porras, A.M. (2001). *Diseño Estadístico de Experimentos, Análisis de la Varianza y Temas Relacionados: Tratamiento Informático mediante SPSS (2ª Edición)*, Ed. Proyecto Sur. Granada (España).

Lara Porras, A.M. (2002). *Estadística para Ciencias Biológicas y Ciencias Ambientales. Problemas y Exámenes resueltos (2ª Edición)*, Ed. Proyecto Sur. Granada (España).

Lara Porras, A.M. et al. (2007). *Técnicas Estadísticas aplicadas en Biología con SPSS*, Ed: Proyecto Sur de Ediciones, S.L.

Lara Porras, A.M. et al. (2008). *Guía interactiva de autoaprendizaje de SPSS versión 2.1 (CD-ROM)*, Ed Proyecto Sur Granada (España). (<http://www.ugr.es/local/bioestad>)

Marqués, Pere (2001). *El impacto de las Tic en la Universidad*. <http://www.pangea.org/peremarques/>

Recursos didácticos para el aprendizaje de la probabilidad: Una experiencia virtual

Lara-Porras, A.M.¹, Román-Montoya, Y.²

Departamento de Estadística e Investigación Operativa. Facultad de Ciencias. Campus de Fuentenueva, Granada

alara@ugr.es, yroman@ugr.es

Resumen

En los últimos años, el aprendizaje virtual se ha convertido en un importante enfoque educativo con destacadas particularidades exclusivas, tanto en el rol de los agentes participantes como en el ambiente de aprendizaje. La virtualidad se ha incorporado en los escenarios educativos como respuesta a las exigencias y necesidades de las nuevas generaciones.

En este proceso de cambio en el que estamos inmersos formulamos este proyecto cuyo principal objetivo consiste en difundir el uso de la utilización de tecnologías para la formación en Probabilidad a través de Internet y capacitar a los alumnos en la utilización de estas técnicas. La importancia de nuestro trabajo radica en la forma en que incorporamos las TICs a los procesos de enseñanza/aprendizaje, redimensionando su utilización no sólo en aspectos de aplicación exclusivamente informáticos, sino también pedagógicos. El alumno aprende el concepto teórico de Probabilidad contrastando los diferentes resultados prácticos que le ofrecen las distintas Webs con las que trabajamos.

Nuestra propuesta metodológica se ha centrado en la asignatura Bioestadística, de carácter troncal con 5.5 créditos, de primer curso de la Licenciatura en Biología desarrollada en la Facultad de Ciencias de la Universidad de Granada, introduciendo a los alumnos en el Cálculo de Probabilidades a través de distintos entornos Web.

La experiencia del curso anterior al aplicar este proyecto en el aula nos permite afirmar que este sistema de trabajo supone un enriquecimiento general en la formación del alumno desde un punto de vista cognitivo, motriz y afectivo-emocional. La amplia variedad de actividades que lo componen ofrecen al alumnado un abanico de posibilidades con un pilar básico: *aprender divirtiéndose*.

Palabras clave: Probabilidad, enseñanza virtual

1. Introducción

El proceso de adaptación de los créditos europeos al sistema de créditos actual ha motivado una importante serie de cambios en el desarrollo de la labor docente. Estos cambios han venido incrementados con la incorporación de las TICs en el aula tanto para la búsqueda de información como para el procesamiento y posterior aplicación de las técnicas analizadas. Nuestra experiencia confirma que la motivación de los alumnos se ve incrementada cuando el docente propone trabajos en equipo basados en la utilización de estas herramientas internautas.

Nuestra propuesta metodológica está basada en proyectos. Intentamos focalizar y dirigir a través de preguntas, problemas o proyectos el aprendizaje de los alumnos, animándoles a trabajar con los principales conceptos de la Probabilidad. Para ello proponemos la realización de actividades que permitan al alumno familiarizarse con las técnicas de investigación, búsqueda y selección de información. El debate, análisis y síntesis de las conclusiones permiten que los conceptos se conviertan en herramientas de aprendizaje más que en objeto de aprendizaje. El objetivo de este trabajo es guiar al alumno, a través de diversas webs centradas en los contenidos indicados (Adell, J. (2004) y Marquès Graells P. (2000))

2. Probabilidad vía Web

La presentación de los contenidos en clase, no sólo es justificada como parte del programa y herramienta de trabajo sino que cada uno de los bloques que constituyen el programa de la asignatura son situados históricamente. De esta forma, comenzamos esta materia hablando a los alumnos de su contexto histórico: El origen del Cálculo de Probabilidades, hace más de 40000 años con los juegos de azar; el desarrollo del análisis matemático de los juegos de azar durante los siglos XVI y XVII y el inicio de la historia de la probabilidad en el siglo XVII cuando Fermat y Pascal tratan de resolver algunos problemas relacionados con los juegos de azar.

A partir de esta información, proponemos a los alumnos un primer trabajo sobre la historia de las Matemáticas. Para ello deben formar equipos de 4-6 personas. En el caso que nos ocupa, el contexto histórico tiene siempre muy buena acogida por parte de los alumnos. Para realizar este trabajo, al igual que para cualquiera de los propuestos a lo largo del curso, orientamos a nuestros alumnos mediante la utilización de diversas páginas Web. En este caso, señalamos:

http://descartes.cnice.mec.es/materiales_didacticos/Historia/historia_indice.html

Esta página, que muestra la evolución histórica del pensamiento matemático, pretende recoger los principales problemas matemáticos ordenados cronológicamente.

Los objetivos que perseguimos con la realización de este trabajo son varios:

- Familiarizar al alumnado con los principales personajes matemáticos y relacionarlos con su época.
- Explicar el porqué y para qué de los temas que forman parte de la asignatura de Estadística.
- Divulgar una matemática que surgió para resolver problemas prácticos, que no nació exacta sino que, en muchos casos, se sirvió de aproximaciones, más o menos justificadas, y que evoluciona hacia la exactitud gracias a la genialidad, el trabajo constante y la colaboración, de muchos hombres interesados en ampliar los límites del conocimiento.

Tras la introducción histórica, presentamos a los alumnos los distintos enfoques de la Probabilidad. Iniciamos este apartado utilizando la figura 1 (obtenida de la red). La presentación de los distintos enfoques de la Probabilidad se hace y se discute en el aula,

pero fundamentalmente es desarrollada por los alumnos. Para ello y dada la extensión de documentación de este tipo en Internet, proponemos a nuestros alumnos trabajar con las siguientes páginas Web:

http://descartes.cnice.mec.es/materiales_didacticos/Azar_y_Probabilidad_jpr/comenzando.htm

En esta página, mediante juegos, distinguimos los experimentos aleatorios de los deterministas y nos introducimos en los diversos conceptos que forman parte de nuestro Tema “Probabilidad”: suceso elemental, espacio muestral,...

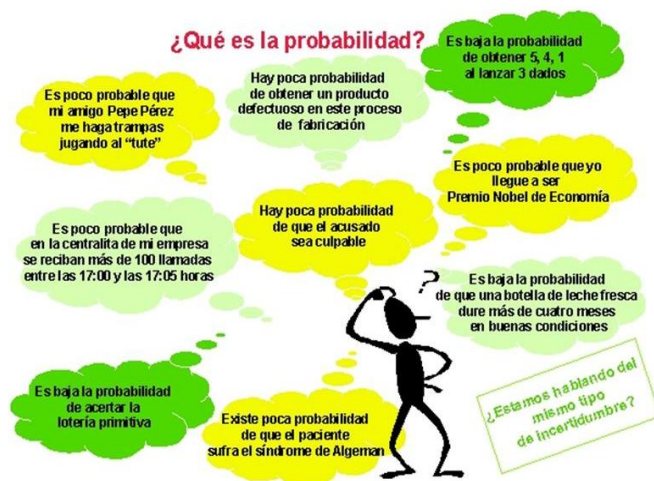


Figura 1: Enfoques de la Probabilidad

Seguimos afianzando conceptos en la siguiente página

http://descartes.cnice.mec.es/materiales_didacticos/Azar_y_Probabilidad_jpr/lasmonedas.htm; donde estudiamos el suceso imposible y el suceso seguro y llegamos a la definición clásica de probabilidad.

En:

http://descartes.cnice.mec.es/materiales_didacticos/Azar_y_Probabilidad_jpr/maquinaria.htm afianzamos los conceptos adquiridos utilizando un nuevo juego: la máquina. Además podemos ver una curiosa relación entre la probabilidad y el número π debida al conde de Buffon, en la dirección http://es.wikipedia.org/wiki/Aguja_de_Buffon.

La Web:

http://descartes.cnice.mec.es/materiales_didacticos/Azar_y_Probabilidad_jpr/elpozo.htm, propone un nuevo juego basado en el lanzamiento de dos dados. Nos familiarizamos así con el concepto de espacio muestral y la ocurrencia de un suceso.

A través de la información disponible en la red, también es posible obtener resúmenes apropiados de lo estudiado, basta observar la página:

http://descartes.cnice.mec.es/materiales_didacticos/Azar_y_Probabilidad_jpr/recuerda.htm

También hacemos repaso, con diversos juegos, de Experimentos aleatorios

http://descartes.cnice.mec.es/materiales_didacticos/Sucesos_aleatorios/sucesos_aleatorios_1.htm y operaciones con sucesos en

http://descartes.cnice.mec.es/materiales_didacticos/Sucesos_aleatorios/sucesos_aleatorios_2.htm

La siguiente Web es una recopilación de todo lo visto ampliándolo con nuevos ejemplos y juegos

http://descartes.cnice.mec.es/materiales_didacticos/Azar_y_probabilidad/index.htm. En esta página de nuevo se observa un primer acercamiento a los conceptos de azar y probabilidad. En ella se introducen los conceptos elementales de la teoría de la probabilidad (suceso, frecuencia, etc.).

Partiendo de ejemplos prácticos se van definiendo los distintos conceptos y se enuncian las reglas y leyes básicas de la probabilidad (regla de Laplace y ley de los grandes números). La página termina con los primeros conceptos de probabilidad de sucesos compuestos atendiendo a los casos de unión e intersección de sucesos.

En la siguiente página se estudia la probabilidad condicionada, el teorema de la probabilidad total y el teorema de Bayes.

http://descartes.cnice.mec.es/materiales_didacticos/probabilidad_condicionada/index.htm.

Los objetivos de esta Web son:

- Comprobar la dependencia e independencia de sucesos.
- Aplicar el teorema de la probabilidad total y el teorema de Bayes para el cálculo de probabilidades sencillas.

Como apoyo, utilizamos las páginas:

http://descartes.cnice.mec.es/materiales_didacticos/probabilidad_condicionada/probabilidad_cond_jam.htm

http://descartes.cnice.mec.es/materiales_didacticos/probabilidad_condicionada/probabilidad_total_jam.htm

http://descartes.cnice.mec.es/materiales_didacticos/probabilidad_condicionada/probabilidad_bayes_jam.htm

<http://ciberconta.unizar.es/leccion/probabil/INICIO.HTML>

Hablaremos de incertidumbre y sus grados, relacionándolos de manera intuitiva con los enfoques más tradicionales para asignar probabilidades. Responderemos a preguntas como ¿Qué es la probabilidad? ¿Cuáles son las propiedades de la probabilidad? ¿Qué dicen los teoremas de la probabilidad total y de Bayes? La utilización de estas páginas permite familiarizar al lector con experiencias de la vida cotidiana en las que interviene el azar; comprender los enfoques de la probabilidad más usuales así como sus peculiaridades, ventajas e inconvenientes; conocer la axiomática de la probabilidad formulada por Kolmogorov; manejar el lenguaje de la probabilidad, sus propiedades y aplicarlo a problemas concretos y entender los teoremas de la probabilidad producto, la probabilidad total y el de Bayes.

Terminaremos con una serie de juegos: el jugador audaz y el jugador constante

http://descartes.cnice.mec.es/materiales_didacticos/JugAudaz/JugadorAudaz.htm

http://descartes.cnice.mec.es/materiales_didacticos/JugAudaz/JugadorConstante.htm

El alumno comprobará que aunque se utilicen estrategias distintas las probabilidades no cambian; la única diferencia estriba en la duración media del juego. Este es un tema que trataremos en la página siguiente:

http://descartes.cnice.mec.es/materiales_didacticos/JugAudaz/PromJug.htm

Terminando con una simulación mediante tablas:

http://descartes.cnice.mec.es/materiales_didacticos/JugAudaz/SimulTablas.htm

Las herramientas Web que hemos utilizado, son gráficas fundamentalmente incluyendo animaciones que facilitan la comprensión de las construcciones gráficas e incorporando actividades que permiten la manipulación de las escenas. En todas se pueden modificar los datos iniciales de los problemas y actuar directamente sobre la gráfica.

El alumno trabaja directamente con el concepto, entiende su origen y su necesidad. Ve experimentalmente como la teoría matemática de la Probabilidad constituye el fundamento de las aplicaciones estadísticas tanto en la investigación social como en la toma de decisiones. La necesidad de sortear la incertidumbre nos lleva a estudiar y aplicar la teoría de la Probabilidad. El éxito en la toma de decisiones, implica la capacidad de tratar sistemáticamente con la incertidumbre mediante evaluaciones y aplicaciones de métodos estadísticos. De este modo, queda completamente justificado y analizado el Cálculo de Probabilidades, como el estudio de métodos de análisis del comportamiento de fenómenos aleatorios.

Una vez sentadas las bases, establecemos el paralelismo existente entre la Estadística Descriptiva y el Cálculo de Probabilidades: las frecuencias relativas se corresponden con el ahora recién incorporado concepto de Probabilidad, las variables se corresponden con los conceptos de variables y vectores aleatorios, la media con la esperanza, etc. y, a partir de aquí, estamos ya en condiciones de introducirnos en el campo de la Inferencia Estadística.

4. Conclusiones

Dentro del proceso de adaptación al sistema ECTS, pretendemos fomentar el protagonismo del estudiante en el ciclo formativo. El estudiante, además de la adquisición de conocimientos, ha de desarrollar una serie de capacidades y competencias prefijadas a partir de la elaboración de actividades académicas dirigidas por el profesor.

Cuando diseñamos este trabajo nuestro objetivo fundamental se centró en el hecho de que los alumnos conozcan y utilicen las TICs como herramienta educativa (Mendaña y González (2004)). Es decir, que trabajen otro modo, más llamativo y estimulante, de adquirir conceptos de la Estadística. Con este sistema de trabajo hemos conseguido una valoración muy positiva del uso del ordenador y de Internet como herramienta poderosa en el aprendizaje de la Estadística. Nuestros alumnos aprenden a utilizar la red como herramienta de trabajo en el estudio. El trabajo en equipo cobra un papel relevante gracias al contraste de ideas y la puesta en práctica de las estrategias consensuadas y desarrolladas conjuntamente en el aula (Ruiz y López (2004)). Además, nuestros alumnos aprenden a pensar de manera lógica y razonada, exponiendo coherentemente sus argumentos como paso previo y necesario para la resolución correcta de un problema.

Los resultados hablan por sí mismos; se aprecia una considerable mejora cuantitativa en las calificaciones finales gracias al mayor interés que los alumnos muestran por la asignatura. El nivel de asistencia y participación se ha incrementado en un 75% y además se observa una creciente y continua demanda por parte del alumnado de nuevos ejercicios y tareas en el desarrollo del temario.

Bibliografía

Adell, J. (2004). Internet en el aula: las WebQuest. Edutec, *Revista Electrónica de Tecnología Educativa*, **17** http://www.uib.es/depart/gte/edutec-e/revelec17/adell_16a.html.

Marquès Graells P. (2000). *Impacto de las tic en educación: funciones y limitaciones* <http://dewey.uab.es/pmarques/si.htm>

Mendaña, C. y González, B. (2004). *El papel de las WebQuest como herramienta para el aprendizaje del alumno en la nueva sociedad del conocimiento*. <http://www.somece.org.mx/virtual2004/ponencias/contenidos/CuervoCristina.htm>. III Simposio Virtual de Computación en la Educación.

http://descartes.cnice.mec.es/indice_ud.php

Ruiz Madrid, N. y López Fernández, O. (2004) El portafolio europeo de lenguas y la sociedad de la digitalización: Una metodología innovadora para la enseñanza-aprendizaje de lenguas y su adaptación al entorno digital. *Revista Latinoamericana de Tecnología Educativa*, **3(1)**, 385-394.

Análisis Exploratorios de Hábitos de Alimentación y Actividad Física en Escolares de la Ciudad de Chilpancingo del Estado de Guerrero México

Vivian Sistachs Vega¹, Lucio Díaz González²

¹ Universidad de la Habana. Cuba. yivian@matcom.uh.cu

² Universidad Autónoma de Guerrero. México. luciodiaz1@yahoo.com.mx

1. Introducción

La obesidad es una enfermedad caracterizada por un incremento en la grasa corporal, que se acompaña de morbimortalidad elevada. Entre los factores asociados al desarrollo de la obesidad se encuentran los factores genéticos, una dieta inadecuada y el sedentarismo, estos últimos característicos de buena parte de la población actual. En un estudio realizado por el Instituto Nacional de Nutrición se encontró que la prevalencia de obesidad en niños y adultos ha aumentado cerca del 50 % en los últimos 10 años, por lo que la Organización Mundial de la Salud (OMS) considera a la obesidad como una verdadera enfermedad epidémica. Los hábitos de alimentación asociados a la actividad física juegan un papel importante en el desarrollo de esta enfermedad.

En la actualidad este problema es muy común en todos los países desarrollados y el aspecto económico de los últimos años se ha convertido en un factor que ha contribuido al estilo de vida y los hábitos alimenticios de la sociedad en general, pareciéndose esta a países industrializados, es decir, una alimentación desbalanceada que se refleja en el estado nutricional de los niños, ya que ellos consumen lo que sus padres les proporcionan y muchas veces no es lo más adecuado.

México se ha caracterizado por tener una dieta rica, en cereales, leguminosas y carnes, así como una gran variedad de verduras y frutas con los cuales se puede preparar un platillo que proporcione todo lo que nuestro cuerpo necesita diariamente, pero debido a la prisa con que vivimos hoy esto se está perdiendo y en la actualidad se tiene una cultura de comida rápida con alto valor energético pero pobre en nutrimentos esenciales que nos lleva a una mala nutrición.

Por lo tanto, a pesar de que nuestra población se consideraba tradicionalmente desnutrida por una alimentación deficiente en calidad y cantidad, desde hace algunos años se ha visto el problema del sobre peso y obesidad, con mayor frecuencia en la población adulta e infantil. Estudios recientes realizados en nuestro país muestran cambios importantes en el estado de nutrición de la población Mexicana. De acuerdo con el XII Censo General de Población y Vivienda (2), 24 millones 157 mil 894 mexicanos tienen problemas de sobrepeso y obesidad, es decir 44.1% del total. Cifra que parece estar por encima de la registrada en la población de Estados Unidos. Según las diferentes regiones, la frecuencia de obesidad es mayor en la región norte del país y en la Ciudad de México comparada con prevalencia del 19.9% en la zona rural, generalmente un área pobre en recursos económicos.

El mexicano destina el 30 % de su ingreso a la alimentación y de este porcentaje el 10% se destina a refrescos. México es el segundo consumidor de refrescos en el mundo, sólo después de Estados Unidos. El 60 % de las familias incluyen el refresco en la dieta habitual dentro de los primeros diez productos de consumo, esto ha venido sustituyendo nuestras tradicionales aguas frescas de frutas (4).

Respecto al estilo de vida, el gran auge de la televisión y de los juegos de video han sido otro factor de gran influencia para el sedentarismo. Hernández et al (5) en 1999 estudiaron a 7112 niños mexicanos entre 9 y 16 años de edad y observaron una clara asociación entre la obesidad, la actividad física y el tiempo empleado en ver televisión: aquellos que veían más tiempo la televisión presentaban mayor riesgo de desarrollar obesidad. En México se lee muy poco, pero en contraposición se ve la televisión. En este sentido, la televisión no sólo modifica conductas y hábitos, sino que promueve el consumo de ciertos productos. De los anuncios publicitarios de televisión destinados a alimentos, el 85% están destinados a la promoción de refrescos, pastelitos y frituras.

La infancia es la etapa en que se desarrollan los hábitos alimenticios y el estilo de vida. El objetivo de este trabajo es realizar un análisis exploratorio para mejorar el nivel de conocimiento sobre los hábitos de alimentación, actividad física y el problema de sobrepeso en los niños de escuelas primarias en Chilpancingo, Gro. México; esto permitiría la implementación de medidas para modificar los hábitos y estilo de vida de esta población.

2. Metodología

El estudio es de tipo transversal. La población de estudio son los escolares de 7 escuelas de la Ciudad de Chilpancingo, Gro. México, que debido a que este es un estudio descriptivo transversal no se utilizó ningún esquema de muestreo y se tomaron todos aquellos niños cuyos padres permitieron su participación en este proyecto, dando un total de 700 estudiantes con el 50% de niños y niñas.

Para conocer sobre la alimentación y el sedentarismo de los niños, se aplicó un cuestionario a los padres, este está dividido en 5 partes: Información Básica, Datos socio-demográficos, Consumo de Alimentos, Actividad física y Medidas Antropométricas. En la parte de consumo de alimentos está dividida en alimentos de alto riesgo, bajo riesgo y protectores.

Para obtener información sobre el estado de salud de los niños se utilizó el IMC, con 4 categorizaciones: Bajo Peso, Normal, Sobrepeso y Obesidad. Para obtener información sobre aspectos socio demográfico y los hábitos de alimentación se utilizó Análisis de Componentes Principales y Análisis factorial de correspondencia simple respectivamente.

3. Resultados

La población bajo estudio está compuesta por 700 escolares con igual porcentaje de niños que de niñas. El porcentaje por escuela se encuentra entre el 8% y 21% conservándose dentro de cada una de ellas el número de escolares varones y mujeres. Respecto al estado nutricional de los escolares, encontramos que un 46.1% tiene bajo peso y un 8% tiene problemas de sobrepeso y obesidad.

Aspecto Social Demográfico

Componente 1: Relación entre el elevado nivel de escolaridad de los padres, así como el ingreso personal de cada uno y su ingreso general asociado a un elevado gasto en el mes en alimentación.

Componente 2: Cantidad de personas que habitan la casa y el alto gasto en el consumo en Kg de azúcares y de aceite.

Componente 3: Llevar mucho tiempo residiendo en el lugar en una casa que no es propia y con pocos ingresos de otro tipo.

Actividad Física Desarrollada.

Componente 1: Lo primero que caracteriza esta actividad es el hecho de que se dedica mucho tiempo a sesiones de televisión, videojuegos y tiempo que usa computadora.

Componente 2: El otro aspecto importante es el hecho de dedicar poco tiempo a la sesión de utilizar la computadora y si mucho tiempo a realizar actividades deportivas fuera de la escuela.

Aspectos Antropométricos

Componente 1: Padres con una talla y peso grande que tiene hijos de talla y peso también grande y que cuando nacieron tenían un peso elevado y un índice de masa corporal (IMC) grande.

Componente 2: Niños con talla y peso pequeños, a pesar de que sus padres tienen estas medidas elevadas esos niños tienen el IMC bajo.

Consumo de Alimentos

Tabla 1: Alimentos de alto riesgo

| Consumo/Alimento | Carne | Carbohidratos | Postres |
|------------------|---|---|--|
| Consume Siempre | Ninguna | Tortilla de maíz, Tacos | Chocolate en polvo con azúcar |
| Consume Mensual | Chorizo Salchicha Pollo y otras aves fritas | Pan dulce Papas fritas Tamal | Néctar de frutas enlatada Gelatina de agua Nieve de agua Hojuelas de maíz |
| Consume Semanal | Res Cerdo Atún | Tortilla de harina de trigo Hot-cakes Galletas con chispas de chocolate | Chocolate de tablilla Fruta en almibar Helado de crema Refresco |
| No Consume | Morongu Ubre Queso chihuahua | Galleta sándwich rellena Galleta a base de mantequilla | Ate o Fruta cristalizada Piloncillo |

Fuente: Encuesta realizado a escolares de 7 escuelas en Chilpancingo, Gro. México.

Tabla 2: Alimentos Protectores

| Consumo/Alimento | Fibra | Frutas | Verduras |
|-------------------------|---|--|----------------------------------|
| Consumo Siempre | Frijol, Soya | Plátano | Ninguna |
| Consumo Mensual | Medias noches Elote Papa hervida Avena | Aguacate Piña Otra fruta | Jícama Lechuga |
| Consumo Semanal | Arroz Pasta para sopa Bolillo con migajón Maíz palomero Masa de nixtamal | Manzana Guayaba Papaya Sandía Mango | Tomate Otras verduras |
| No Consume | Bolillo sin migajón Pan integral Cereal de salvado Camote | Yogurt natural | Zanahoria |

Fuente: Encuesta realizado a escolares de 7 escuelas en Chilpancingo, Gro. México.

Los resultados en el consumo de alimentos de alto riesgo, bajo riesgo y protectores, existe evidencia de que los alumnos consumen mucho los productos con carbohidratos y azúcares, así como un alto consumo de maíz, frijol, soya, plátano y leche entera, estos resultados tienen mucho sentido ya que los productos con un alto porcentaje de consumo son parte de la canasta básicas mexicana, sin embargo es necesario que estos productos se consuman de manera balanceada y esto se logra teniendo buenos hábitos de alimentación.

En cuanto a los hábitos de alimentación encontramos que el 98% de los niños realiza dos comidas al día y un porcentaje menor 80% cenan, también podemos decir que 52% tiene el mal hábito de comer cuando estudian o ven televisión.

4. Conclusión

La proporción de sobrepeso y obesidad en niños escolares de las escuelas de Chilpancingo, Gro. (8.0%) es mucho menor que el bajo peso (46.1%). En cuanto al aspecto socio demográfico podemos decir que existe una relación entre el nivel escolar de los padres y su ingreso, esto permite un gasto mensual elevado en la alimentación. También podemos ver evidencia en el consumo de alimentos de alto riesgo aquellos que contienen carbohidratos y azúcares; en los alimentos de bajo riesgo existe un consumo en el huevo frito, la leche entera, pan y la pasta para hacer sopa, esto tiene sentido ya que son productos de la canasta básica mexicana, con respecto alimentos protectores podemos destacar que los niños no consumen fibra, zanahoria y sobre todo verduras, pero existe evidencia de un alto consumo de frijol y soya, así como el plátano.

En cuanto a la actividad física se destaca que los estudiantes dedican mucho tiempo a realizar actividades sedentarias, esto quiere decir que se mantienen mucho tiempo sentados sin hacer actividades físicas.

Respecto a los hábitos de alimentación podemos decir que los niños realizan al menos dos comidas al día (almuerzo y comida), sin embargo tienen el mal hábito de comer cuando ven televisión o estudian.

REFERENCIAS

- Norma Oficial Mexicana NOM-174-SSA1-1998, para el Manejo Integral de la Obesidad.* Publicación Oficial 1998/010198
- Instituto Nacional de la Nutrición. Encuestas Nutricionales en México. México: División de Nutrición, 1963.
- Encuesta Nacional de Salud y Nutrición (2006)
- Instituto Nacional de Estadística, Geografía e Informática. Encuesta Nacional de Ingresos y Gastos de los Hogares, México: INEGI, 2000.
- Hernández B, Gortmakers SL, Goldiestz GA. Association of obesity with physical activities, television programs and other forms of video viewing among children in Mexico City. *Int Obesity* 1999; 23: 845-54.
- Johnson (2000), *Multivariate analysis*, ed. John Wiley
- Jambu (1990), *EXploratory Data Analysis*, ed. John Wiley
- Levi S, Rivera DJ, Ramírez SC. Sobrepeso y obesidad en población escolar mexicana. Libro de resúmenes del IX Congreso Nacional de Investigación en Salud Pública, 2001: 34-35.
- Manual software de SPSS 10.0
- Manual de R version 2.8.0 (2008).

Análisis Estadístico de Datos Textuales

M. Luisa Hdez. Maldonado, Claudio R. Castro López
Facultad de Estadística e Informática. UV

Resumen

Los investigadores frecuentemente se plantean el estudio de datos de tipo textual provenientes de documentos históricos, resúmenes de artículos, traducciones de libros a distintos idiomas, contenidos de bases científicas, corpus voluminosos, encuestas que contienen preguntas abiertas o de respuesta libre, etc. El tratamiento clásico para el estudio de este tipo de datos textuales es costoso desde el punto de vista de personal empleado, especialistas y tiempo; ante esta situación, la Estadística Textual ha integrado distintas técnicas y métodos de análisis que facilitan el proceso de la exploración de este tipo de datos. El cartel presenta la metodología que se recomienda en el tratamiento de datos provenientes de una pregunta abierta.

Introducción

Los problemas dependen de la disciplina a la que pertenecen y de los intereses del investigador: por ejemplo, en corpus voluminosos puede interesar conocer las tendencias en el lenguaje, comparar cronológicamente el uso del vocabulario, comparar estilos, hacer una búsqueda documental, etc.; el estudio de una bases de datos histórica sobre pacientes de un hospital, el estudio puede centrarse en conocer características de los pacientes y la sintomatología que presentaban; los discursos políticos, artículos, crónicas deportivas, pueden ser datos portadores de información. En las respuestas a preguntas abiertas el tratamiento se puede enriquecer con la información complementaria obtenida de las respuestas del resto de las preguntas del cuestionario.

Este cartel se centra en el Análisis Estadístico de Datos Textuales (AEDT) provenientes de preguntas abiertas contenidas en un cuestionario. El AEDT contempla procedimientos que permiten hacer un análisis y depuración del vocabulario hasta obtener un glosario de formas (palabras y segmentos repetidos) con los que se va a trabajar y a realizar algún tipo de análisis estadístico a partir de los resultados de la selección hecha. Utiliza técnicas que se enmarcan dentro del análisis de exploración multivariada, los resultados se pueden enriquecen con un análisis de clasificación (Lebart y Salem 1988, Bécue 1991). Los resultados del estudio permiten al investigador, identificar rasgos que se presentan en los individuos, semejanzas entre ellos y asociaciones con su vocabulario. La riqueza y diversidad de los resultados obtenidos le permiten al investigador comprobar hipótesis, señalar hallazgos y/o plantear nuevas investigaciones.

Análisis de datos textuales

Los métodos de la estadística textual han surgido de la necesidad de hacer estudios cuantitativos de los textos literarios, y del llamado análisis de datos, que es la tendencia de la estadística moderna (Zipf, G.K., 1935). Posteriormente (Yule, 1944; Guiraud, 1960; Muller, 1968) la estadística lexical se enfrentó con problemas planteados por estilistas preocupados por el estudio comparativo del vocabulario de los grandes autores: comparación de la riqueza del vocabulario, análisis de la evolución del vocabulario de un mismo autor, etc. El uso de las computadoras facilitaron estos procesos de análisis.

Actualmente, el soporte informático con que se cuenta permite que el análisis de datos textuales se aplique a todo tipo de textos y se pueda complementar con otros métodos léxico-métricos y técnicas clásicas como el estudio del vocabulario (Lebart y Salem 1988; Bécue 1991). Evidentemente, en cada caso, se toman en cuenta las perspectivas de investigación propias de las distintas disciplinas que entran en relación con los textos (historia, pedagogía, sociología, psicología, etc.). El análisis de clasificación proporciona grupos de individuos homogéneos en cuanto a las palabras utilizadas en sus respuestas o bien, grupos de palabras frecuentemente utilizadas por los mismos individuos.

Las preguntas abiertas

Las preguntas de respuesta libre son muy ricas en información y sin embargo, no son muy frecuentes en las encuestas, debido a que las respuestas obtenidas presentan gran dificultad para su tratamiento estadístico posterior. Frecuentemente se utilizan en encuestas pequeñas o en la fase preparatoria, principalmente con el objetivo de obtener información que permita formular adecuadamente preguntas cerradas.

El desarrollo actual de la estadística textual permite el tratamiento sistemático de la información obtenida en una encuesta en la que se incluyen preguntas abiertas o preguntas de respuesta libre, incluso en muestras de gran tamaño. La estadística textual considera las respuestas libres en su forma original, es decir, las respuestas completas emitidas por los entrevistados a la pregunta abierta. De esta manera tiene en cuenta las palabras y sus contextos. Además utiliza la información sobre las características de los encuestados, recogidas en preguntas cerradas.

La estadística textual proporciona métodos de análisis que permiten realizar un estudio de la encuesta desde una perspectiva multivariante. Utiliza toda la información disponible del encuestado sin realizar una reducción a priori, de este modo se evitan deformaciones, sesgos y pérdidas de información. Se puede analizar un elevado número de respuestas para cada preguntas abiertas contenida en el cuestionario.

Análisis textual a preguntas abiertas

La estadística textual se apoya en técnicas estadísticas desarrolladas por la escuela francesa de análisis de datos. La base de este método consiste, a grosso modo, en utilizar procedimientos que implican contar las ocurrencias de las unidades verbales básicas (generalmente palabras) y realizar algún tipo de análisis estadístico a partir de los resultados de dichos recuentos. A continuación se describen los elementos fundamentales del análisis textual.

Formas gráficas y segmentos repetidos

La unidad de base será la forma gráfica definida como una sucesión de caracteres no delimitadores (normalmente letras) comprendidos entre dos delimitadores (blanco y signos de puntuación) (Lebart y Salem 1988). Ejemplos de formas gráficas “idea”, “conozco”, “buena”. El conjunto de formas de un texto constituye su vocabulario.

Una vez conocido el vocabulario de texto, interesa conocer los perfiles de las frecuencias de las formas gráficas y especialmente las diferencias entre ellos. El perfil léxico es un vector cuyos componentes son las frecuencias de cada una de las formas utilizadas por un individuo o un grupo de individuos.

Los segmentos repetidos se definen como una secuencia de dos o más palabras, no separadas por un delimitador de secuencia, que aparecen más de una vez en un corpus de datos textuales. Ejemplos de segmentos repetidos son “es muy buena”, “sin comentarios”. Son unidades más largas, compuestas por varias formas, dotadas de un sentido propio y que no siempre se deduce de las formas que entran en su composición, si éstas se consideran aisladamente.

Para el tratamiento estadístico se asigna a cada forma gráfica diferente un número de orden, el mismo siempre que aparece. La numeración da lugar a la creación de un diccionario de formas propio de cada análisis. Todo tratamiento estadístico para las formas gráficas es aplicable a los segmentos repetidos.

Tablas léxicas

Las respuestas libres pueden ser numeradas de manera completamente automática por el analista y, a partir del diccionario de formas, construir tablas rectangulares que serán analizadas mediante las técnicas multivariantes descriptivas, como el Análisis de Correspondencias y el Análisis de Clasificación (Benzécri y otros, 1973, 1981; Abascal y Grande, 1989; Escofier y Pagès, 1992; Lebart, Morineau y Piron, 1995).

Estas tablas recogen la distribución de palabras entre las distintas respuestas individuales o entre grupos de individuos. Los grupos se definen de acuerdo con alguna característica de los entrevistados, recogida en sus respuestas a una pregunta cerrada. Las tablas utilizadas son:

- La tabla léxica no agregada T (individuos x palabras). Es una matriz de k filas, una por cada respuesta, y un número de columnas igual al de formas gráficas utilizadas por el conjunto de individuos. En la intersección de la fila i y la columna j , figura la frecuencia con que aparece la forma j en la respuesta del individuo i . Esta tabla recoge los perfiles léxicos de los individuos.
- La tabla léxica agregada (palabras x grupos de sujetos). Es la tabla de contingencia cuyo término general recoge el número de veces que la forma i ha sido utilizada en la respuesta libre, por el conjunto de individuos que eligen la modalidad j en la respuesta cerrada. A partir de esta tabla se pueden comparar los perfiles léxicos de los segmentos de población definidos. Se puede obtener una tabla léxica agregada para cada pregunta cerrada, y comparar los perfiles léxicos de las diferentes categorías de la población.

Técnicas descriptivas de análisis de datos multivariantes.

El análisis de datos textuales se apoya en los métodos de análisis estadístico de grandes tablas de datos, concretamente en el análisis de correspondencias y en el análisis de clasificación. El análisis de correspondencias es un método descriptivo clásico para el estudio de tablas de rectangulares. Describe la asociación entre las dos variables categóricas correspondientes a la tabla. Se puede ver también, como la búsqueda de la mejor representación simultánea de las modalidades de las dos variables categóricas y estudia las relaciones eventuales existentes entre ellas; proporciona una visión gráfica, simplificada, de la información que contiene poniendo en evidencia las diferencias entre los distintos perfiles léxicos. El análisis puede incluir de manera ilustrativa variables de tipo nominal y numérico.

Las representaciones permiten situar a los individuos en un espacio determinado por las formas, permitiendo comparar los perfiles de sus respuestas, localizando aquellos más semejantes y las formas causantes de estas semejanzas. Así mismo, se detectan los individuos más diferentes explicando sus causas. La representación de las formas permite estudiar las proximidades entre aquellas formas que son utilizadas simultáneamente por los mismos individuos, es decir, estudiar los contextos, ya que la proximidad entre dos formas es mayor cuando aparecen en la misma respuesta.

Si se analiza la tabla léxica agregada se pueden estudiar las posiciones relativas de diferentes grupos de individuos en relación al vocabulario empleado y caracterizarlos por las palabras utilizadas.

Con este tipo de análisis, la presencia de palabras herramienta (preposiciones, artículos, conjunciones, pronombres, etc.) está en principio justificada si estas palabras caracterizan a algunas categorías, o su posición próxima puede ser interesante para interpretarlas. Si por el contrario, su reparto es aleatorio y se aproximan o se sitúan en la parte central de la gráfica, interesa eliminarlas. Por otra parte, la presencia próxima de varias formas de un mismo verbo, puede constituir una herramienta de validación y justificar la lematización (declarar equivalentes las formas gráficas que tienen la misma raíz y un significado equivalente); aunque, se ha comprobado que esta etapa no es indispensable y puede incluso presentar problemas (Lebart, 1986) porque frecuentemente son utilizadas en diferentes contextos y pueden tener diferentes significados.

Análisis de Clasificación

Continuando con el tratamiento a las respuestas de una pregunta abierta, se determinan las clases que clasifiquen a la totalidad de la muestra encuestada. La clasificación normalmente se hace a partir de las primeras diez coordenadas factoriales. El corte jerárquico se realiza cuando el índice de distancia realiza un salto importante, así, se elige un número determinado de clases. Como resultado de la clasificación se describe la caracterización de cada una de las clases.

Conclusiones

- Los métodos de análisis de estadística textual conforman un instrumento muy útil para el análisis de preguntas de respuesta libre.
- Los distintos métodos utilizados: análisis de correspondencias de la tabla léxica, análisis de correspondencia de la tabla léxica agregada y el análisis de clasificación de las formas, pueden proporcionar resultados poderosos en el estudio de los individuos.

REFERENCIAS

- Abascal, E. y Grande, I. (1989) Métodos multivariantes para la Investigación Comercial. Barcelona: Ariel.
- Bécue, M. (1991) Análisis de datos textuales. Métodos estadísticos y algoritmos. CISIA.
- Benzécri JP y col. (1973) La taxonomie ,VolI. L'Analyse des Correspondances, Vol II. Paris: Dunod.
- Benzécri JP y col.(1981) Practique de l'Analyse des Données, tome III, Linguistique et Lexicologie . Paris: Dunod.
- Escofier, B. y Pagès, J.,(1992) Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación. Universidad del País Vasco.
- Guiraud, P. (1960). Problèmes et Méthodes de la Statistique Linguistique. Paris: PUF.
- Lebart, L.(1986) Analyse statistique des réponses libres dans les enquêtes par sondage. Revue Française du Marketing, 109.
- Lebart, L; Salem, A. (1988) Analyse statistique de données textuelles. Questions ouvertes et lexicométrie . Bordas.
- Lebart, L.; Morineau, A. y Piron, M. (1995). Statistique exploratoire multidimensionnelle. Paris: Dunod.
- Muller, C. (1968). Initiation a la Statistique Linguistique. Paris: Larousse.
- Zipf, G.K. (1935). The Psychobiology of Language, an Introduction to Dynamic Philology, Boston: Houghton-Mifflin.

A Consideration for Constructing Critical Regions for Non-inferiority Tests

Dr. Félix Almendra-Arao

UPIITA del Instituto Politécnico Nacional. falmendra@ipn.mx

Abstract

In non-inferiority tests the concept of licit sample space is of great importance since if sample space is not restricted to licit sample space can lead to invalid inferences. In this work we consider non-inferiority tests that use the difference between two proportions as the measure of dissimilarity. We give examples of tests for which the critical regions corresponding to the sample space and licit sample space are equal as well as an example of a test for which such critical regions are different.

Keywords: Non-inferiority, Licit Sample Space, Statistical Tests, Critical Regions.

1. Introduction

Non-inferiority tests are statistical procedures used to demonstrate that a new treatment's effectiveness is not significantly inferior to a standard treatment. Recently these tests have acquired much importance; however, there are several related aspects that need special attention, particularly, the aspect concerning the notion of licit sample space, which is a restriction of the sample space. The concept of licit sample space was developed by Martin and Herranz, 2002, 2004a and 2004b, who emphasized its importance. However, we consider it relevant to illustrate this concept in more detail, revealing the importance of the concept of licit sample space as a restriction of the sample space. When the sample space is not restricted to licit sample space, erroneous inferences may be obtained in some cases.

In this paper we present some non-inferiority tests for which critical regions corresponding to sample space and licit sample space are the same. Also, we give an example of a non-inferiority test for which the critical regions are different. The goal of this work is to call attention to the need to use the concept of licit sample space in non-inferiority tests, in general terms, and specifically, in the generalized likelihood ratio test.

2. Non-Inferiority Tests for the Difference Between Proportions

Let X_1 and X_2 be two independent binomial random variables with parameters (n_1, p_1) and (n_2, p_2) , respectively, where it is assumed that p_1 and p_2 represent the true response probabilities of the standard and new treatments, respectively. For the hypothesis testing problem

$$H_0 : p_1 - p_2 \geq d_0 \quad \text{vs} \quad H_a : p_1 - p_2 < d_0 \quad (1)$$

where d_0 is a known constant, referred to as margin of non-inferiority.

In the specific case of hypothesis testing problems of type (1), the sample space (SS) is $\mathcal{X} = \{(x_1, x_2) \in \{0, \dots, n_1\} \times \{0, \dots, n_2\}\}$ and the parameter space is $\Theta = \{(p_1, p_2) \in [0, 1]^2\}$.

Given a hypothesis testing problem (1), the licit sample space (LSS) is defined as

$$\mathcal{X}' = \left\{ (x_1, x_2) \in \mathcal{X} : \hat{p}_1 - \hat{p}_2 < d_0 \right\} \quad \text{where} \quad \hat{p}_i = X_i / n_i \quad \text{is the maximum likelihood}$$

estimator of p_i for $i=1,2$, according to Martin and Herranz, 2002, 2004a and 2004b. The reason for this restriction is that, otherwise, the test would not be significant; for example, if $\hat{p}_1 - \hat{p}_2 \geq d_0$, then is inferentially illogical to conclude that $p_1 - p_2 < d_0$ (see Martin and Herranz, 2004).

We will use R_T to denote the critical region (CR) for a test T for the hypothesis testing problem (1). The licit critical region (LCR), that is, the critical region obtained when the SS is restricted to LSS, denoted by R_T^L is given by $R_T^L = \left\{ (x_1, x_2) \in R_T : \hat{p}_1 - \hat{p}_2 < d_0 \right\}$.

3. Some Tests with Identical Critical and Licit Critical Regions

When we consider statistical tests of the form

$$T(X_1, X_2) = \frac{\hat{p}_1 - \hat{p}_2 - d_0}{\hat{\sigma}} \quad (2)$$

for the hypothesis testing problem (1), where $\hat{\sigma}$ is an estimator of the standard deviation of $\hat{p}_1 - \hat{p}_2$, the CR corresponding to a test T is of the form $R_T = \{ (x_1, x_2) \in \chi : T(x_1, x_2) \leq t_0 \}$, whereas the LCR is of the form $R_T^L = \{ (x_1, x_2) \in \chi' : T(x_1, x_2) \leq t_0 \}$, where t_0 is a known constant.

For tests that use type (2) statistics, the following proposition proves that CR and LCR are equal.

Proposition.

When type (2) statistics are used for the hypothesis testing problem (1), CR and LCR are equal. That is, if T is a type (2) statistic and $t_0 < 0$, then $R_T^L = R_T$.

Proof.

By definition, $R_T^L \subseteq R_T$. Suppose that $R_T - R_T^L \neq \emptyset$ then there exists $(x_1, x_2) \in R_T - R_T^L$, as

$(x_1, x_2) \in R_T$, then $\frac{\hat{p}_1 - \hat{p}_2 - d_0}{\hat{\sigma}} \leq t_0$. Moreover, as $(x_1, x_2) \notin R_T^L$, then $\hat{p}_1 - \hat{p}_2 \geq d_0$;

therefore $\frac{\hat{p}_1 - \hat{p}_2 - d_0}{\hat{\sigma}} \leq 0$ and $\frac{\hat{p}_1 - \hat{p}_2 - d_0}{\hat{\sigma}} > 0$. This is a contradiction, therefore, it must

be true that $R_T - R_T^L = \emptyset$. Thus, *CR and LCR are equal.* \square

There are several type (2) statistics. Two well known are established in Blackwelder, 1982 and Miettinen and Nurminen, 1985 and Farrington and Manning, 1990. As both have normal standard asymptotic distribution, it follows that $t_0 = -z_\alpha < 0$, where z_α is the upper percentile α of the normal standard distribution. That is, $P(Z > z_\alpha) = \alpha$; Z is a normal standard random variable.

The test given in Blacwelder, 1982 uses

$$\hat{\sigma}_1 = \left(\frac{\hat{p}_1 \left(1 - \hat{p}_1 \right)}{n_1} + \frac{\hat{p}_2 \left(1 - \hat{p}_2 \right)}{n_2} \right)^{1/2}$$

to estimate the standard deviation of $\hat{p}_1 - \hat{p}_2$, whereas the test given in Miettinen and Nurminen, 1985 and Farrington and Manning, 1990 uses

$$\hat{\sigma}_2 = \left(\frac{\check{p}_1 \left(1 - \check{p}_1 \right)}{n_1} + \frac{\check{p}_2 \left(1 - \check{p}_2 \right)}{n_2} \right)^{1/2}$$

to estimate the standard deviation of $\hat{p}_1 - \hat{p}_2$, where \check{p}_i is the maximum likelihood estimator of p_i , restricted under the null hypothesis. We denote the respective tests by T_1 and T_2 . Thus, by the previous proposition, critical regions and licit critical regions are the same. That is, $R_{T_i}^L = R_{T_i}$ for $i=1,2$.

Of course, the above proposition is also valid for the exact version of both tests, T_1 and T_2 , when the critical constant (t_0) is negative, as required by this proposition.

4. A Test with Unequal Critical and Licit Critical Regions

Martin and Herranz. 2002 give data for which for the generalized likelihood ratio test critical region and licit critical region not necessarily are equal.

The statistics for the generalized likelihood ratio test is

$$\lambda(X_1, X_2) = \frac{\sup_{\Theta_0} L(d | (X_1, X_2))}{\sup_{\Theta} L(d | (X_1, X_2))}$$

in which $d = p_1 - p_2$, the asymptotic distribution of $-2 \ln \lambda$ is $\frac{1}{2} + \frac{1}{2} F_{\chi_1^2}$, where $F_{\chi_1^2}$ denotes the accumulated distribution of a chi-square random variable with one degree of freedom.

The critical region for the corresponding asymptotic test is

$$R_T(\alpha) = \left\{ (x_1, x_2) : -2 \ln \lambda(x_1, x_2) > \chi_{1-2\alpha}^2(1) \right\}$$

where $\chi_{1-2\alpha}^2(1)$ is the upper $1-2\alpha$ percentile of the chi-square distribution with one degree of freedom; in other words, $P(\chi_1^2 > \chi_{1-2\alpha}^2(1)) = 1 - 2\alpha$.

In figure 1, CR (R_T) and LCR (R_T^L) regions are illustrated for $n=30$, $\alpha=0.05$, $d_0=0.10$. CR has many more points than LCR; each point that belongs to R_T but not to R_T^L is a point that will give an invalid inference. In other words, such points have no significance.

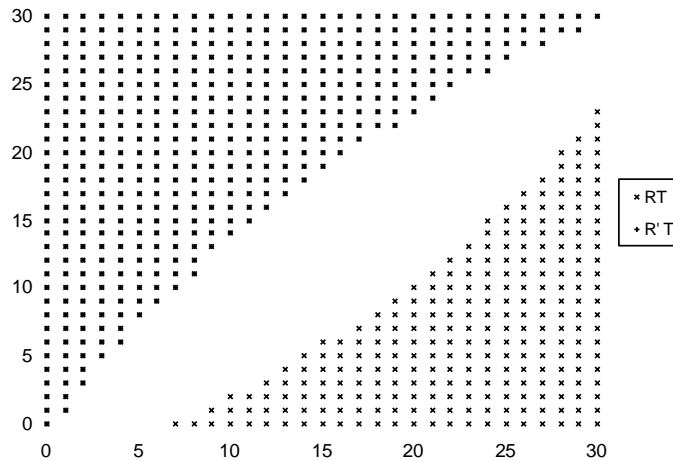


Figure 1. CR (R_T) and LCR (R_T^L) for $n = 30$, $\alpha = 0.05$, $d_0 = 0.10$.

Numerically, we verified that for sample sizes $30 \leq n_1 = n_2 \leq 100$, R_T^L satisfies the Barnard convexity condition, whereas R_T does not. According to Röhmel J, and Mansmann, 1999, if a critical region satisfies the Barnard convexity condition, calculating the significance level is reduced considerably. In fact, although considering LCR provides that collateral advantage, it is of vital importance, because it allows computing significance levels for this test in a practical way.

As, in general, R_T^L has fewer points than R_T , then the corresponding power function is also lower.

We showed that inferentially it is incorrect not to consider the LSS in constructing critical regions for tests. Martin and Herranz, 2002 give some sample points that have significance when the LSS is not considered but are not significant when the LSS is considered. It might be thought that if the difference between CR and LCR is only few points, it is not transcendental from a practical point of view. From an inferential logical perspective, however, this is a mistake even when the difference is only one point.

Moreover, for the asymptotic generalized likelihood ratio test, when $\alpha=0.05$ and $d_0=0.10$, we quantify the measure in which the critical region increases when the LSS is not considered in its construction.

For sample sizes $30 \leq n_1 = n_2 = n \leq 100$, the cardinalities of CR and LCR were calculated. The percentage in which LCR increases CR non-significantly is high for analyzed sample sizes. This leads to a large error in all calculations, in particular, in calculations of significance levels and power.

4. Conclusions

We give examples of non-inferiority statistical procedures for which it is not important whether the LSS is used or not. Moreover, we give an example of another statistical procedure, which is affected negatively when the LSS is not considered.

We exemplified several tests whose CR and LCR are equal and one test in which they are different, the generalized likelihood ratio test.

In constructing critical regions for non-inferiority tests, the LSS must be taken into consideration, and therefore, LCR, R_T^L , must be constructed instead of R_T , unless for some reason it is known that both regions are equal, as for example, in the case of the proposition presented in this study.

REFERENCES

- Blackwelder, W. C. (1982). "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials*. 3, 345-353.
- Farrington C, Manning G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine* 9, 1447-1454.
- Martin, A. and Herranz, I. (2002). Equivalence testing for binomial random variables: which test to use? *The American Statistician* 56(3), 253-254, Letter to the editor.
- Martin, A. and Herranz, I. (2004a). Asymptotical test on the equivalence, substantial difference and non-inferiority problems with two proportions. *Biometrical Journal*. 46, 305-319.
- Martin, A. and Herranz, I. (2004b). Exact unconditional non-classics tests on the difference of two proportions. *Computational Statistics & Data Analysis*. 45,373-388.
- Miettinen O, Nurminen M. (1985). Comparative analysis of two rates. *Statistics in Medicine* 4, 213-226.
- Röhmel J, and Mansmann U. (1999). Unconditional nonasymptotic one sided tests for independent binomial proportions when the interest lies in showing noninferiority and or superiority. *Biometrical Journal* 2, 149-170.

La Explicación del Funcionamiento de un Sistema Mediante la Modelización de Ecuaciones Estructurales

M^a Elena Vicente¹, J. Antonio Castro², M^a Purificación Vicente³, M^a Purificación Galindo⁴

¹ Universidad Pontificia Salamanca. España. canaryavg@hotmail.com

² Universidad de Salamanca. España. juancastrap@gmail.com

³ Universidad de Salamanca. España. purivg@usal.es

⁴ Universidad de Salamanca. España. pgalindo@usal.es

Introducción

Aristóteles estableció que al conocimiento se llega por sus causas. No es de extrañar, por tanto, que desde la Filosofía de la Ciencia se siga postulando que conocer es explicar (Fodor, 1980). Por eso, tal vez la mejor explicación sea aquella que se ofrece desde las conexiones entre las causas y sus efectos. El presente trabajo trata de mostrar cómo se puede conseguir la explicación del funcionamiento de un sistema mediante la metodología de los modelos de ecuaciones estructurales (Structural Equation Modeling), aunque respetando la flexibilidad de la metodología que pone de manifiesto la importancia de la toma de decisiones del científico por encima de los propios planteamientos metodológicos. Este trabajo plantea la necesidad de aceptar la relatividad de la metodología frente al conocimiento y el manejo teórico del investigador. La metodología es el instrumento, y, como tal, nunca puede dejarse la inferencia en manos de los avatares metodológicos. Es el científico quien construye la ciencia; nunca lo será el cálculo.

Método

Sujetos: Se utilizó una muestra de 552 sujetos universitarios con edades comprendidas entre 18 y 24 años (\bar{x} = 21 años; s = 2); el 64.9% eran mujeres y el 35.1% varones. La distribución por carreras fue la siguiente: Ciencias Sociales, el 12.3%; Ciencias de la Salud, el 12.6%; Ciencia y Tecnología, el 16.8%; Gestión y Leyes, el 14.2%; Diplomaturas, el 15.3%; y Otras, el 4.1%.

Procedimiento: Mediante el programa informático AMOS, v.5 (Arbuckle, 1999), se procedió a la modelización de las relaciones causa-efecto, mediante path analysis, entre las 16 variables o ítems que configuraban los factores “autoestima emocional” y “autoestima cognitivo-comportamental” de la escala CASVI, la cual ha sido diseñada para evaluar inteligencia emocional en estudiantes universitarios (véase, Vicente, 2007 y Vicente, Castro, Vicente y Galindo, 2008). Se construyó un modelo inicial en el que 10 de las variables funcionaron como endógenas y las 6 restantes intervinieron como exógenas; el modelo mostró un alto nivel de ajuste. A la vista de la bondad de ajuste del modelo, se procedió a la construcción de un modelo alternativo con 7 variables endógenas y 9 variables exógenas; el modelo mostró, también, un alto nivel de bondad de ajuste. A continuación, se construyó un tercer modelo tomando como variables endógenas aquellas que repetían esta función en ambos modelos, y como variables exógenas aquellas que también repetían dicha función en los dos modelos. El resultado fue un modelo reducido, en el que intervinieron 6 variables endógenas y 3 variables exógenas, que mostró, asimismo, un alto nivel de bondad de ajuste.

Resultados

Modelo inicial: El modelo inicial, construido según la metodología del prototipo (véase Castro, 2002), quedó constituido por 6 variables exógenas puras (causas), por 4 variables endógenas puras (efectos) y por 6 variables intermediarias (que funcionan como efecto de otras variables y, a la vez, como causa de otras distintas) (Ilustración 1).

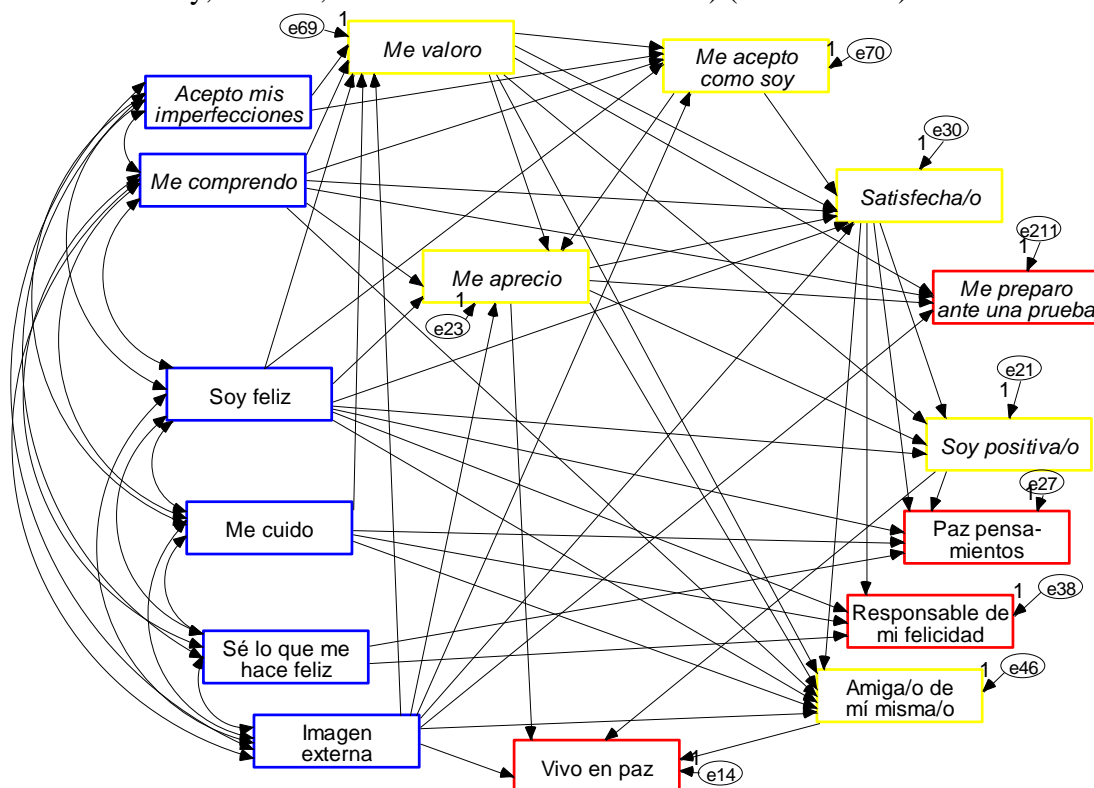


Ilustración 1. Modelo inicial. Las variables en caja azul son exógenas puras; las que aparecen en caja roja son endógenas puras, y las que se ofrecen en caja amarilla son variables intermediarias, o endo-exógenas. Las variables nombradas en cursiva pertenecen al ámbito cognitivo-comportamental; el resto pertenece a la dimensión afectiva o emocional de la autoestima (Vicente, Castro, Vicente y Galindo, 2008).

Modelo alternativo: Se construyó un segundo modelo, alternativo, tratando de hacer funcionar las variables en sentido contrario al primer modelo, es decir, aquellas que eran exógenas en el primer modelo, se trataron como endógenas, y viceversa. El resultado aparece en la Ilustración 2. Como puede verse, son, ahora, 9 las variables exógenas puras, 3 las variables endógenas puras, y 4 las variables intermediarias. Es de destacar la trama amplia de covarianzas que se estableció entre las variables exógenas puras, lo que da una idea de la gran consistencia o coherencia del sistema causal o exógeno de este modelo (35 covarianzas, todas ellas altamente significativas ($p < 0,000$), frente a las 15 covarianzas del primer modelo, todas ellas también altamente significativas ($p < 0,000$)).

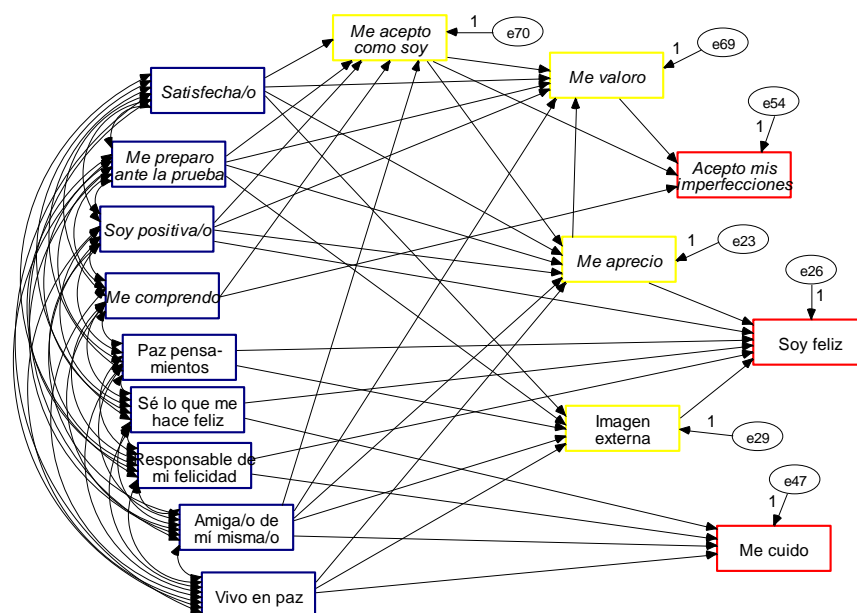


Ilustración 2. Modelo alternativo. Las variables en caja azul son exógenas puras; las que aparecen en caja roja son endógenas puras, y las que se ofrecen en caja amarilla son variables intermediarias, o endo-exógenas. Las variables nombradas en cursiva pertenecen al ámbito cognitivo-comportamental; el resto pertenece a la dimensión afectiva o emocional de la autoestima (Vicente, Castro, Vicente y Galindo, 2008).

Modelo reducido: A la vista de las coincidencias y diferencias que ofrecen los dos modelos (véase la tabla 1), se construyó un tercer modelo, el modelo reducido (Ilustración 3), para el que se utilizaron aquellas variables que cumplían la misma función en los dos modelos anteriores, es decir, en ambos modelos aparecían como exógenas, o endógenas, o intermediarias (exo-endógenas).

Tabla 1. Función de cada una de las variables en los 3 modelos. Ex = exógena; En = endógena; I = intermediaria. Las variables nombradas en cursiva pertenecen al ámbito cognitivo-comportamental; el resto pertenece a la dimensión afectiva o emocional de la autoestima (Vicente, 2007; Vicente, Castro, Vicente y Galindo, 2008).

| Variable | Modelo Inicial | Modelo Alternativo | Modelo Reducido |
|---|----------------|--------------------|-----------------|
| Vivo en paz | En | Ex | |
| <i>Soy positiva/o</i> | I | Ex | En |
| <i>Me aprecio</i> | I | I | I |
| Soy feliz | Ex | En | |
| Paz pensamientos | En | Ex | |
| Imagen externa | Ex | I | Ex |
| <i>Satisfecha/o</i> | I | Ex | I |
| Responsable de mi felicidad | En | Ex | |
| <i>Me comprendo</i> | Ex | Ex | Ex |
| Amiga/o de mí misma/o | I | Ex | En |
| Me cuido | Ex | En | |
| Sé lo que me hace feliz | Ex | Ex | Ex |
| <i>Acepto mis imperfecciones</i> | Ex | En | |
| <i>Me valoro</i> | I | I | I |
| <i>Me acepto como soy</i> | I | I | I |
| <i>Me preparo ante una prueba (211)</i> | En | Ex | |

El modelo que aparece en la Ilustración 3 muestra 3 variables exógenas puras, junto a 2 variables endógenas puras y 4 variables intermeditarias.

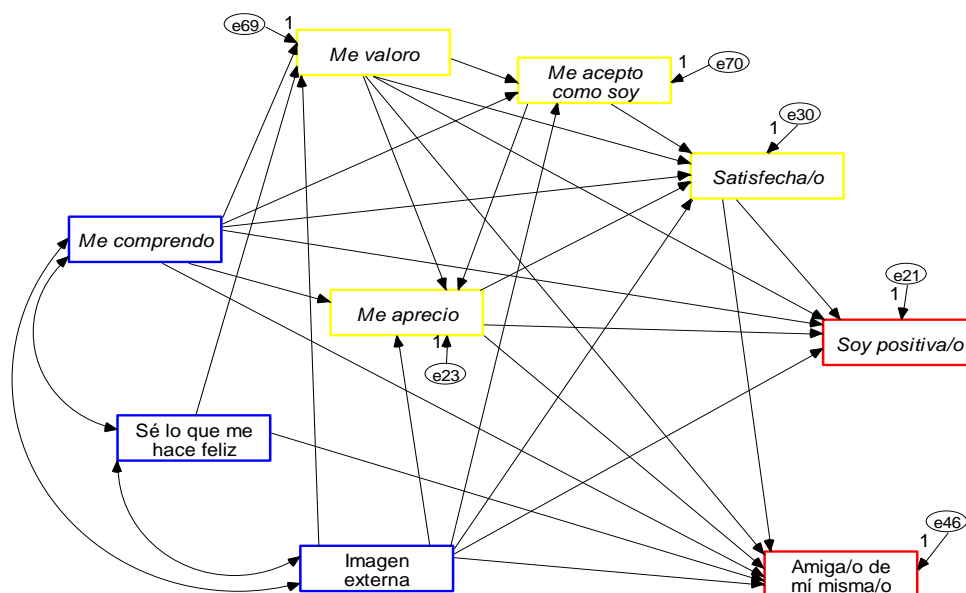


Ilustración 3. Modelo reducido. Las variables en caja azul son exógenas puras; las que aparecen en caja roja son endógenas puras, y las que se ofrecen en caja amarilla son variables intermeditarias, o endo-exógenas. Las variables nombradas en cursiva pertenecen al ámbito cognitivo-comportamental; el resto pertenece a la dimensión afectiva o emocional de la autoestima.

Los tres modelos ofrecen una excelente bondad de ajuste (tabla 3).

Tabla 3. Bondad de ajuste de los 3 modelos.

| Modelo | NPAR | CMIN | DF | P | CMIN/DF |
|--------------------|-----------------------|---------------------|-----------------------|---------------------|------------|
| Modelo inicial | 80 | 70,767 | 56 | ,088 | 1,264 |
| Modelo alternativo | 87 | 42,257 | 49 | ,741 | ,862 |
| Modelo reducido | 38 | 5,151 | 7 | ,642 | ,736 |
| | RMR | GFI | AGFI | PGFI | |
| Modelo inicial | ,032 | ,984 | ,962 | ,405 | |
| Modelo alternativo | ,021 | ,990 | ,974 | ,357 | |
| Modelo reducido | ,012 | ,998 | ,987 | ,155 | |
| | NFI Delta1 | RFI rho1 | IFI Delta2 | TLI rho2 | CFI |
| Modelo inicial | ,980 | ,958 | ,996 | ,991 | ,996 |
| Modelo alternativo | ,988 | ,971 | 1,002 | 1,005 | 1,000 |
| Modelo reducido | ,998 | ,988 | 1,001 | 1,004 | 1,000 |
| Modelo | RMSEA | LO 90 | HI 90 | PCLOSE | |
| Modelo inicial | ,022 | ,000 | ,036 | 1,000 | |
| Modelo alternativo | ,000 | ,000 | ,021 | 1,000 | |
| Modelo reducido | ,000 | ,000 | ,043 | ,976 | |

Consideraciones Finales

Dado que los tres modelos obtenidos son válidos (su bondad de ajuste es excelente), es evidente que tienen que emerger una serie de preguntas para la discusión, entre las que conviene destacar, al menos, las siguientes: ¿Cuál de ellos es el mejor? ¿Cuál es el valor de la metodología? ¿Cuál es el papel del científico ante esta diversidad? ¿Tiene que ver en estos resultados la posible existencia de una causalidad circular en la explicación científica? ¿Es la parsimonia un criterio de selección del mejor modelo posible? Dado que los modelos pertenecen al ámbito de la Psicología, ¿son transferibles sus implicaciones inferenciales a otros ámbitos de la producción científica? Sin duda, el mejor modelo será el que decida el científico, basándose en la potencia de cálculo que le ofrece esta metodología.

Otra de las ventajas que ofrece esta metodología es el abordaje de la causalidad circular que, si bien en Psicología está, al menos implícitamente, asumida, en otras ciencias parece que todo apunta a la “linealidad” de la causalidad. No obstante, es digno de tener en cuenta que toda ciencia que trate de explicar sistemas adaptativos complejos (Biología, Medicina, etc.) no puede renunciar a abordar este problema de la circularidad.

Sea como fuere, lo que sigue estando claro es que la discusión sobre la explicación científica sigue abierta (entre otras cosas, porque sigue abierta la cuestión de la causalidad), sobre todo desde el momento en que se dispone de tecnología y metodología adecuadas para validar modelos de relaciones causales entre variables. Pero ésta es la ventaja: hoy día, en pocos segundos se puede estimar la validez de un modelo, y con algún tiempo más, se pueden construir modelos alternativos. Lo importante es que el conjunto de modelos obtenidos permite al científico reflexionar y decidir a tenor de sus conocimientos teórico-científicos. La ciencia ha sido, y seguirá siendo, un conjunto de toma de decisiones por parte de los científicos, aunque, eso sí, basándose en instrumentos de cálculo cada vez más sofisticados.

REFERENCIAS

- Arbuckle, J. L. (1999). *Amos 4.0 [Computer software]*. Chicago, IL: SmallWaters.
- Castro, J.A. (2002). *Metodología de la investigación. Diseños*. Salamanca: Amarú.
- Fodor, J. A. (1980). *La explicación psicológica*. Ediciones Cátedra.
- Vicente M. E. (2007). *Creación de una escala para medir inteligencia emocional. Estudio piloto. Tesina Licenciatura*. Universidad Pontificia. Salamanca. España.
- Vicente, M.E., Castro, J.A., Vicente, M.P. y Galindo, M.P. (2008). *Estudio de la validez del constructo inteligencia emocional, en estudiantes universitarios, mediante un modelo de análisis factorial confirmatorio: escala CASVI*. México: XXIII Foro Nacional de Estadística. Universidad Veracruzana.

Modelos para Enfermedades Infecciosas

Vivian Sistachs Vega¹, Lucio Díaz González², Lucio Díaz González³

¹ Universidad de la Habana. Cuba. vivian@matcom.uh.cu

² Universidad Autónoma de Guerrero. México. luciodiaz1@yahoo.com.mx

³ Maestría en Estadística Aplicada UAGro. México. marisol_gb@hotmail.com

1. Introducción

El brote y la propagación de enfermedades ha sido cuestionada y estudiado durante muchos años. La capacidad de hacer predicciones acerca de las enfermedades podría permitir a los científicos a evaluar los planes de vacunación y el aislamiento, esto pueden tener un efecto significativo sobre la tasa de mortalidad de una epidemia. El modelado de las enfermedades infecciosas es una herramienta que se ha utilizado para estudiar los mecanismos por los que se propagan las enfermedades, para predecir el futuro curso de un brote y evaluar estrategias para el control de una epidemia (Daley y Gani, 2005).

El primer científico que sistemáticamente trató de cuantificar las causas de muerte fue John Graunt en su libro *Natural y Políticos Observaciones hechas en el momento de los proyectos de ley de mortalidad*, en 1662. Los proyectos de ley que se estudia las listas de los números y las causas de las muertes publicado semanalmente. Graunt del análisis de las causas de muerte se considera el comienzo de la "teoría de los riesgos competitivos" que, según Daley y Gani (Daley y Gani, 2005, p. 2) es "una teoría que está ahora bien establecida entre los epidemiólogos modernos".

La primera cuenta de la modelación matemática de propagación de la enfermedad se llevó a cabo en 1766 por Daniel Bernoulli. Entrenado como un médico, de Bernoulli creó un modelo matemático para defender la práctica de vacunar contra la viruela (Hethcote, 2000). En los cálculos de este modelo demostró que la vacunación universal contra la viruela podría aumentar la esperanza de vida de 26 años 7 meses a 29 años y 9 meses (de Bernoulli y Soplante, 2004).

A raíz de Bernoulli, otros médicos han contribuido a la epidemiología matemática moderna. Entre las más aclamadas de estos fueron AG McKendrick y WO Kermack, cuyo documento de *Contribución a la Teoría Matemática de las epidemias* se publicó en 1927.

El presente trabajo se presenta la clasificación de los modelos matemáticos para enfermedades infecciosas, se describen de forma detallada algunos modelos típicos y otros esquemas recientes que se utilizan para modelar las enfermedades infecciosas, se muestran los resultados de la aplicación de algunos de estos modelos en enfermedades infecciosas.

2. Los Modelos Matematicos

La construcción de modelos matemáticos es una herramienta que se aplica en Medicina, Física, Biología, Epidemiología, entre otra. Sus objetivos son describir, explicar y predecir fenómenos en dichas áreas. La relevancia de la construcción de los modelos matemáticos para enfermedades infecciosas es evidente, ya que un modelo puede mostrar relaciones que nos se puede ver a simple vista, es posible extraer propiedades y características de las

relaciones entre sus elementos y sobre todo porque no es posible experimentar en la realidad debido al elevado costo, peligroso e imposible.

La función principal de un modelo para enfermedades infecciosas es conocer la propagación de una enfermedad en una determinada población, para prevenir futuras situaciones de dicha enfermedad, determinar la prevalencia e incidencia y apoyar en la toma de decisiones para controlar y/o erradicar las enfermedades.

3. Tipos de Modelos

Existen dos tipos de modelos matemáticos: determinísticos y estocásticos. En un modelo determinístico se pueden controlar los factores que intervienen en el estudio del proceso o fenómeno y por tanto se pueden predecir con exactitud sus resultados. En un modelo estocástico no es posible controlar los factores que intervienen en el estudio del fenómeno y en consecuencia no produce simples resultados únicos. Cada uno de los resultados posibles se genera con una función de probabilidad que le asigna una probabilidad a cada uno de éstos por ejemplo un modelo para predecir el tamaño de una epidemia en una población de N individuos. Para el caso determinístico se proporciona un valor único, C , mientras que el modelo estocástico permite la posibilidad de obtener desde cero hasta N individuos y se asigna una cierta probabilidad a cada uno de estos sucesos.

La diferencia es más grande de lo que parece, ya que en un modelo matemático determinístico en el contexto epidemiológico; un solo sujeto causa una epidemia generalizada, mientras que bajo un modelo estocástico existe la posibilidad de que la epidemia se extinga.

4. Clasificación de los Modelos Matemáticos

Existen varios tipos de modelos que son considerados en el estudio de una epidemia en una población dada. Estos modelos dependen del tipo de enfermedad bajo estudio. Por ejemplo algunas enfermedades confieren permanentemente inmunidad después de la infección mientras otras se manifiestan en el individuo después de un corto llamado periodo latente. Algunos ejemplos de modelos se muestran en la tabla 1.

5. Formas de Transmisión

Transmisión vertical: En algunos casos de enfermedades tal como SIDA y Hepatitis B, es posible para la descendencia de padres infectados ser infectado. Este tipo de transmisión a través de la madre es llamada transmisión vertical.

Transmisión: Enfermedades transmitidas de humano a humano indirectamente, por ejemplo un brote de malaria por de mosquito, se transmiten a través de un vector. En este caso, la infección se transfiere de humano a insecto y el modelo debe incluir a ambas especies.

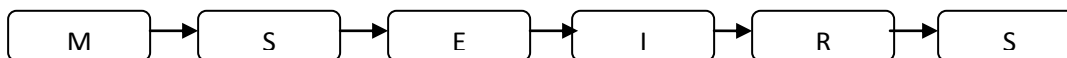
Tabla 1. Modelos para enfermedades infecciosas

| Modelo | Uso | Enfermedad |
|--------|--|-----------------------|
| SI | Describir modelos en los cuales el individuo infectado no se recupera de la enfermedad | SIDA |
| SIS | Representa un modelos donde la población infectada no tiene inmunidad, esto significa que individuos recuperados pueden ser infectados con la enfermedad | Tuberculosis, sífilis |

| Modelo | Uso | Enfermedad |
|--------|---|---|
| SIR | Describir a modelos donde individuos infectados que recuperan de una enfermedad dada tiene inmunidad permanente, mueren o está recuperándose de la población. | Hepatitis, polio, difteria, Cólera, sarampión |
| SIRS | Este modelo corresponde a casos donde individuos se infectan y recuperan perdiendo su inmunidad y se vuelven otra vez susceptibles. | Dengue, influenza, viruela, colera |
| SEIR | Este modelo es utilizado para describir modelos con periodo latente E, donde el individuo es infectado con una particular enfermedad pero no muestra síntomas. | Viruela |
| SEIS | Este modelo tome en consideración el periodo expuesto y latente de la enfermedad. En este modelo, la infección no deja una inmunidad de larga duración por lo tanto, las personas que han recuperado volver a ser susceptibles de nuevo | Gripe |
| MSIR | Estudiar aquellas enfermedades donde el individuo nace con la inmunidad pasiva de su madre. | Tuberculosis, fiebre tifoidea |
| MSEIR | Este modelo se emplea en aquellas enfermedades, con el factor de inmunidad pasiva y un periodo de latencia. | |
| MSEIRS | Este modelo es similar a un MSEIR pero la inmunidad puede ser temporal y los individuos podrían recuperar su susceptibilidad cuando la inmunidad temporal termina. | |

El diagrama general para el modelo MSEIRS se presenta en la figura 1, en este diagrama se presentan la diferentes etapas por las que un individuo puede pasar al contraer una enfermedad en la cual se utilice un MSEIRS. M=Inmunidad pasiva, S=Susceptibles, E=Expuestos, I=Infectados, R=Recuperados.

Figura 1. Diagrama de transferencia para modelos MSEIRS



6. Aplicaciones

En la tabla 2 se muestran algunas aplicaciones de los modelos para enfermedades infecciosas, en esta tabla se presentan las herramientas utilizadas para el estudio de dicha enfermedad.

| Aplicación | Herramientas | Modelo empleado | Referencia |
|---|---|-----------------|------------|
| Casos de sarampión | - Gibbs Sampling - Montecarlo dentro de Metropolis | SIR | 2007 |
| Estudio de Influenza en Alemania | - Ecuaciones diferenciales | SIR | 2005 |
| Análisis de series de tiempo para conteos de enfermedades infecciosas | Métodos MCMC | SIR | 2005 |
| Estudio de un modelo dinámico para el análisis y simulación de las enfermedades transmisibles en una población repartida en tres grupos de población; susceptibles, infectados e inmunes. | Simulación | SIR | 2006 |
| Estructura global de un modelo | Simulación | SI | 2003 |

| Aplicación | Herramientas | Modelo empleado | Referencia |
|---|-------------------------------------|-----------------|------------|
| espacial de enfermedades infecciosas | | | |
| Estudio de dos grupos de datos: viruela y enfermedad respiratoria | Salto reversible del algoritmo MCMC | SEIR | 2007 |

7. Conclusion

El empleo de los modelos matemáticos para enfermedades infecciosas ha crecido significativamente y cada vez más enfermedades son modeladas mediante estos modelos permitiendo conocer su comportamiento bajo determinadas condiciones, permitiendo su control y en algunos casos su erradicación.

Las herramientas utilizadas para la aplicación de estos modelos depende de las características de la enfermedad y de los aspectos que se incorporen en ellos, tanto que una misma enfermedad puede ser modelada por varios modelos. Asimismo, hay que resaltar que existen modelos epidemiológicos sencillos que modelan algunas enfermedades de manera adecuada y no es necesario ser un experto en la materia para hacer uso apropiado de ellos, pero existen modelos muy complejos que no son de fácil comprensión para el experto en la materia y para hacer uso de ellos es necesario contar con herramientas matemáticas y computacionales más sofisticadas.

Por otro lado es necesario mencionar que no existe una metodología sistemática que permita la modelación de enfermedades infecciosas de manera única, sin embargo es importante que los profesionales de salud y epidemiología tengan conocimiento de estos modelos, para tomar medidas de control y erradicación de las enfermedades.

REFERENCIAS

- Beardmore I., Beardmore R., "The global structure of a spatial model of infectious disease", The royal society, 2002.
- Cooke K. L. "Analysis of an SEIRS epidemic model", Mathematical Biology, No. 35, 1996
- Dargatz C., Georgescu V., Held L., "Stochastic modeling of the spatial spread of influenza in Germany", Sonderforschungsbereich 386, paper 450, 2005
- Green P. J., "Reversible jump Markov chain monte carlo computation and Bayesian model determination", Department of Mathematics Unniversity of Brisol, 1996.
- Hethcote W. H., Tudor W. D., "Integral Equation Models for Endemic Infectious Diseases", Methematical Biology, No. 9, 1980.
- Hethcote W. H. "The Mathematics of Infectious Diseases", Society for Industrial and Applied Mathematcs, Vol. 42, No. 4, 2000
- Held L., Hofmann M., Schmid V., "A two-component model for counts of infectious diseases", Department of Statistics, Ludwin-Maximilians Universitat Munchen, 2005
- Montesinos L. O, Hernández S. C., "Modelos Matemáticos para enfermedades infecciosas", Salud Pública de México, Vol. 49, no.3, 2007
- O'Neil D. P., et al., "Analyses of Infectious disease Data from Household outbreaks by markov chain monte carlo methods", 2005.

Estructura Causa Suficiente en Modelos Gráficos para Enfermedades Complejas con Expresiones Orales

Mabel Brunotto¹, Silvia Mina, Magdalena Bolesina, Silvia Sorokin, Irma Moncunill, Susana Cornejo, Vanesa Buffa, Dante Secchi, Eduardo Piemonte, Evelin Bachmeier, Jerónimo Lazos, María Brezzo, Ana María Zarate, Fabiana Carletto Köber, Alicia Malberti, Silvina Barembaum, Ana Isabel Azcurra, José Luis Barra

Universidad Nacional de Córdoba

¹brunottomabel@gmail.com

Resumen

El *objetivo general* del presente trabajo fue diseñar y evaluar modelos gráficos directed acyclic (DAG) con estructura de causa suficiente para valorar relaciones causales entre un conjunto de variables clínicas médica-odontológica, psicológicas, moléculo-celulares y los fenotipos multifactoriales de enfermedades complejas. Se evaluaron variables binarias que se incorporaron a un modelo DAG con estructura de causa suficiente para cada patología.

Palabras clave: *estructura causa suficiente- enfermedades complejas- DAG*

Introducción

Actualmente, en salud se generan modelos que explican la causalidad de un proceso bio-psico-social que genera una enfermedad. Estos modelos se evalúan, tanto, desde una perspectiva estadística como desde los fundamentos biomédicos subyacentes causantes de los eventos (Brunotto et al. 2009, Clayton 2001, Thornton-Wells et al. 2004).

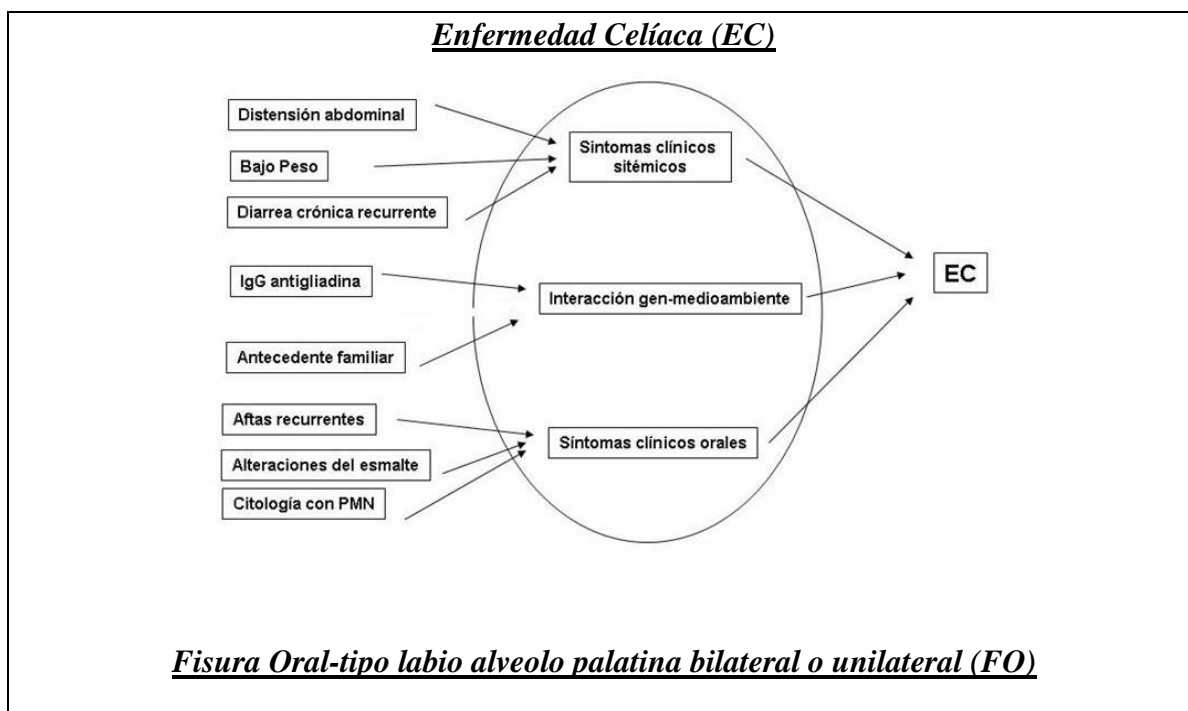
Los modelos gráficos pueden ser interpretados como un conjunto diagramático para la generación de ecuaciones estructurales (Marchetti, 2006), que permiten evaluar el impacto de las variables en la generación de una patología. Estas gráficas consisten en un conjunto de variables y de caminos directos (*paths*) entre ellas, conformando un gráfico directo acíclico, DAG. Este tipo de modelos, que establecen relaciones causa-efecto, permitiría valorar por parte de los profesionales de la salud la verdadera influencia que tiene cada variable, haciendo posible generar programas de prevención en poblaciones de riesgo o proyectos de profundización de los estudios genético-moleculares de las patologías complejas. Enfermedades multifactoriales con expresión de signos y síntomas en cavidad oral, como celiaquía, cáncer oral y fisuras orales, implican una dificultad inherente para su diagnóstico temprano y su adecuado tratamiento en poblaciones de riesgo. Estas patologías presentan particularidades en los factores psicológicos, medioambientales, socioculturales y genéticos que las afectan, pero comparten no seguir las leyes mendelianas de la herencia. Los componentes genéticos, ambientales y psico-sociales interactúan generando una gama de fenotipos clínicos generales y a nivel de cavidad oral que dificulta el diagnóstico presuntivo y/o temprano. Si bien, estas enfermedades no presentan una prevalencia importante a nivel poblacional, producen serios trastornos de vida en el paciente; entre los que se pueden mencionar: procesos inflamatorios leves, formas neoplásicas malignas, alteraciones en la morfología y función de las piezas dentarias, malformaciones anatómicas, entre otras (Mina et al, 2008, Sorokin et al, 2008, Zarate et al, 2007).

El *objetivo general* del presente trabajo fue diseñar y evaluar modelos gráficos directos acíclicos (DAG) con estructura de causa suficiente para valorar relaciones causales entre un conjunto de variables clínicas médica-odontológica, psicológicas, moleculo-celulares y los fenotipos multifactoriales de enfermedades complejas.

Material y Método

Se realizó un estudio observacional transversal en grupos poblacionales con patologías complejas de celiacía, fisura oral y cáncer oral, mediante muestreo no probabilístico. Nuestro equipo interdisciplinario obtuvo el valor de las variables características de tres patologías crónicas complejas: *Enfermedad Celíaca*, *Fisura Oral* y *Cáncer Oral*. Los datos clínicos, genéticos (mutaciones de genes específicos mediante PCR y secuenciación), medioambientales y psico-socioculturales fueron registrados mediante Historia Clínica que contenía una conjunto de datos comunes y las variables características de cada patología. Se evaluaron variables binarias según criterios establecidos por investigaciones previas (Mina et al, 2008, Sorokin et al, 2008, Zarate et al, 2007) que se incorporaron a un modelo DAG con estructura de causa suficiente para cada patología (VanderWeele and [Robins](#), 2007).

Los DAGs para cada patología basados en un modelo teórico sustentado en los conocimientos biológicos-clínicos-psico-sociales de cada patología se muestran en la Figura 1.



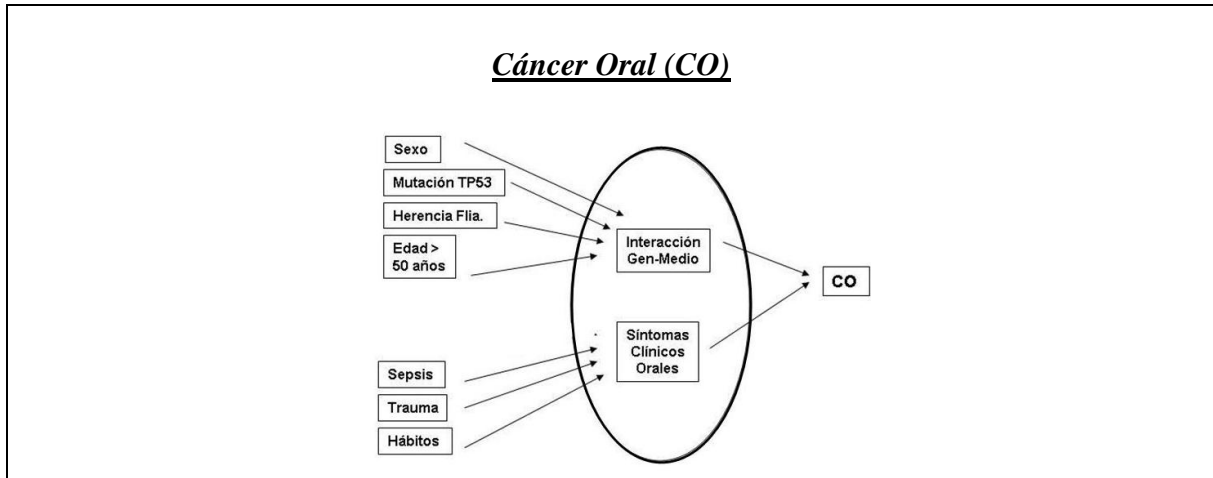
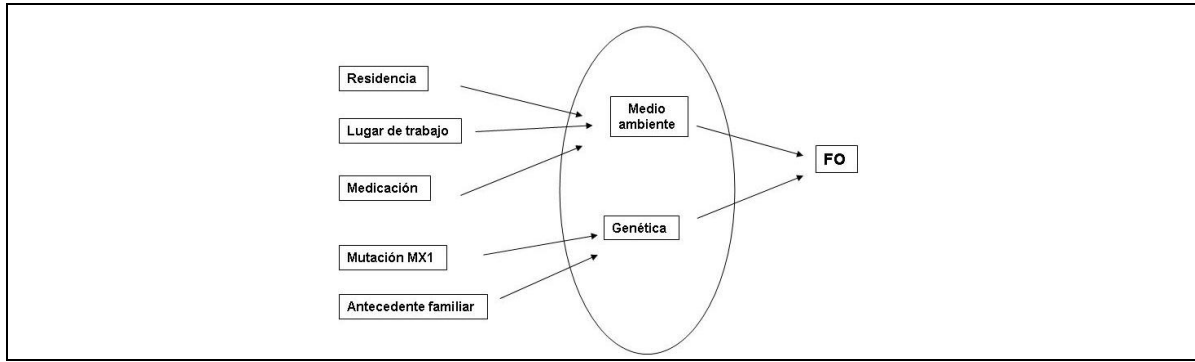


Figura 1. Modelos DAG con estructura causa suficiente.

Resultados

En el análisis de los datos se utilizó el paquete *ggm* del programa R, versión 2.9.0 (www.r-project.org). Se obtuvieron, mediante la función *dag*, las gráficas $D = (V, E)$ donde V es el conjunto de nodos y E es el conjunto de líneas directas (Tabla 1, Figura 2); y sus correspondientes matrices adyacentes $E = [e_{ij}]$. A partir del conjunto de líneas faltantes, que corresponden a los ceros estructurales de la matrix E , se evaluaron las independencias condicionales, llamadas propiedades globales Markov del DAG $f_Y(y) = \prod_{i \in V} f(y_i | y_{pa(i)})$, y que se leen por el criterio de *d-separation* (Marchetti et al, 2006). Se corroboraron, en la mayoría de las variables, las propiedades de *d-separación* de cada gráfica (Tabla 2).

Tabla 1. Sentencias para obtener los modelos DAG para cada patología. EC: enfermedad celíaca; CO: cáncer oral; FO: fisura labio-alveolo-palatina

| Patología | Sentencias, ggm package, función dag |
|-----------|--|
| FO | dag_FO<DAG(gen~MSX1+Ant_Her,Medio~Exp_lab+resid+medic, TF~gen+Medio, order=TRUE) |
| CO | dag_CO<DAG(SCO~trauma+sepsis+hab,Evol_lesion~edad1+sexo+mutación+ant_flia, PATOL~SCO+Evol_lesion, order=TRUE) |
| EC | dag_cel4<DAG(Sign_Sint~Ba_Peso+Dist_Abdo+DAR,sint_oral~Aftas+Esmalte+Citol,Exp_glu~antec_flia+IgG_AGA,grupo~Exp_glu,grupo~Sign_Sint+sint_oral+Ex_glu,order=TRUE) |

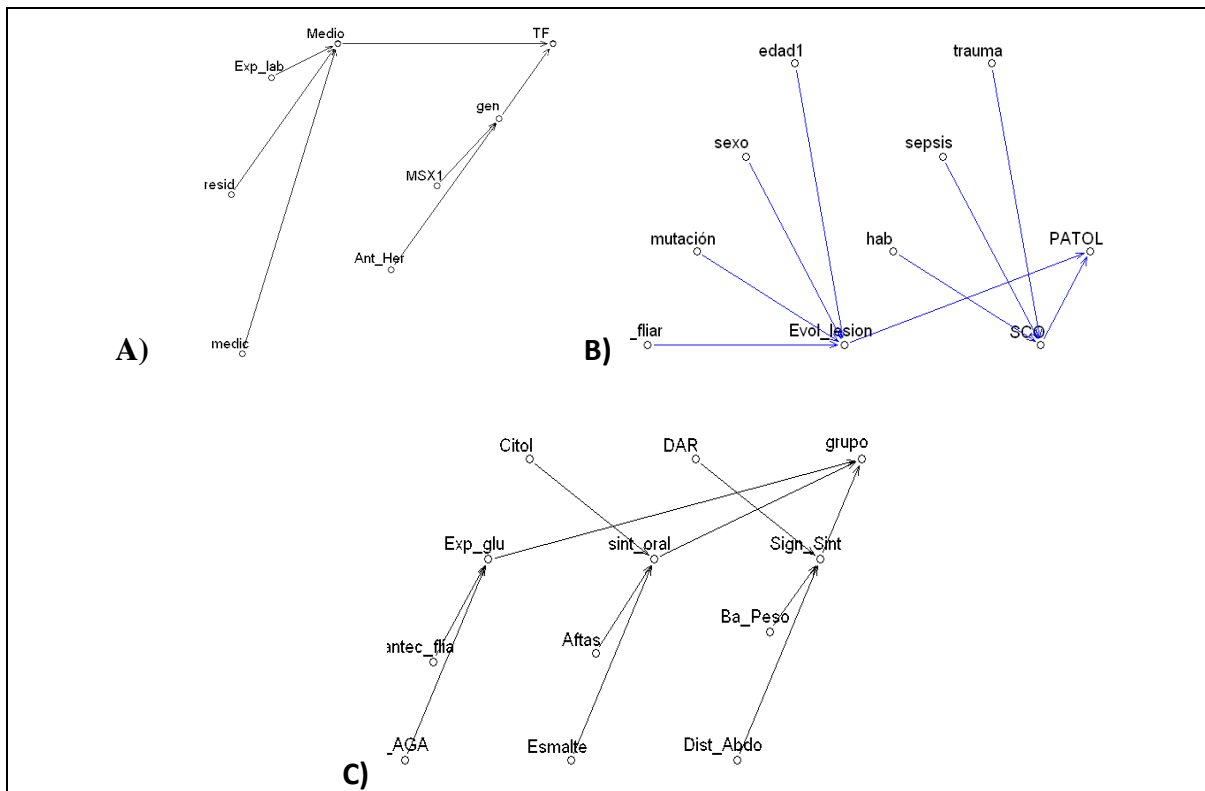


Figura 2. DAGs de cada patología construidos con el paquete ggm del software R.
A) FO; B) CO; C) EC

Tabla 2. Criterio de d-separación para cada patología, estimado con la función dSep del paquete ggm del programa R, versión 2.9.0

| Patología | Independencias condicionales de Markov |
|-----------|---|
| FO | $MSX1 \perp TF gen$ |
| | $Ant_Her \perp TF gen$ |
| | $Exp_lab \perp TF Medio$ |
| | $resid \perp TF Medio$ |
| | $medic \perp TF Medio$ |
| CO | $trauma \perp PATOL SCO$ |
| | $sepsis \perp PATOL SCO$ |
| | $hab \perp PATOL SCO$ |
| | $edad1 \perp PATOL Evol_lesion$ |
| | $ant_fliar \perp PATOL Evol_lesion$ |
| | $sexo \perp PATOL Evol_lesion$ |
| EC | $Ba_Peso \perp grupo Sign_Sint$ |
| | $Dist_Abdo \perp grupo Sign_Sint$ |

| | |
|--|------------------------------------|
| | DAR \perp grupo Sign_Sint |
| | Citol \perp grupo sint_oral |
| | Aftas \perp grupo sint_oral |
| | Esmalte \perp grupo sint_oral |
| | antec_flia \perp grupo Exp_glu |
| | IgG_AGA \perp grupo Exp_glu |

Discusión

Casi todas las independencias condicionales fueron chequeadas, aunque algunas resultaron falsas, como la variable mutación en relación a cáncer oral. Este resultado es razonable dado que el gen que se amplificó y secuenció (TP53), se presenta en sólo un 50% de las lesiones ocasionadas por esta patología (Brunotto et al. 2009).

Este trabajo presenta dos aspectos importantes a considerar:

1. La forma de trabajo interdisciplinaria, de un equipo de profesionales de varias áreas de las ciencias de la salud (biólogos, médicos, psicólogos, odontólogos, bioquímicos), que permite el adecuado abordaje de las patologías complejas;
2. La aplicación de modelos gráficos como una herramienta valiosa para el mejor ajuste del modelo estadístico permitiendo una mejor correspondencia con el modelo de salud real de estas enfermedades.

REFERENCIAS

- Brunotto M, Zarate AM, Barra JL, Malberti A. Graph models for phenotype and genotype association between oral mucosa and submandibular gland tumorigenesis in rat. *J Oral Pathol Med.* 2009 May;38(5):463-9.
- Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; 358: 1356–60.
- Marchetti GM. Independencies induced from a Graphical Markov Model after marginalization and conditioning: the R package ggm. *J Statist Software.* 2006; 15:1–15 (<http://www.jstatsoft.org>).
- Mina S, Azcurra A, Riga C, Dorronsoro S, Brunotto M. Alterations of Buccal Ecosystem in children with celiac disease. *Acta Odontol Latinoam.* 2008;21(2):121-6.
- Sorokin S, Brunotto M, Bolesina M, Moncunill I, Cornejo LS, Buffa V. Prevalencia de tipo y magnitud de fisuras orales desde 1978 a 2005, Facultad de Odontología Universidad Nacional de Córdoba. *Rev Cúspide.* 2008. 17:225.
- Thornton-Wells TA, Moore JH, Haines JL.. Genetics, statistics and human disease: analytical retooling for complexity. *Trends Genet* 2004; 20: 640–7.
- VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am J Epidemiol.* 2007 Nov 1;166(9):1096-104.
- Zarate AM, Secchi D, Brezzo M, Bachmeier E, De Niro I, Garay P, Brunotto M. Cambios genotípicos-fenotípicos como predictor de malignización en lesiones orales en la ciudad de Córdoba, Argentina. *Braz Oral Res* 2007; 21 (Suppl.1):123 (Proceedings of the 24th SBPqO Annual Meeting)

Ajuste, Predicción y Tiempos de Primer Paso en un Modelo Estocástico para el Estudio de Crecimiento del Pez Espada

Román-Román, Patricia¹, Torres-Ruiz, Francisco de Asís²

¹Departamento de Estadística e I.O. Universidad de Granada. proman@ugr.es

²Departamento de Estadística e I.O. Universidad de Granada. fdeasis@ugr.es

Introducción

La curva de crecimiento de Bertalanffy

$$f(t) = f_{\infty} [1 - e^{-k(t-a)}]; \quad t \geq a; k > 0.$$

donde f_{∞} es la cota superior y k es el parámetro *curvatura* que determina la velocidad con la que se alcanza la cota, se emplea comúnmente para modelar crecimiento de animales y, en particular, de peces. Para ello existen modelos determinísticos y estocásticos asociados con esta curva, permitiendo éstos últimos la inclusión de fluctuaciones o perturbaciones que pueden existir en el sistema bajo consideración y que no siempre pueden ser cuantificables o que, incluso, pueden ser desconocidas. Esta curva se emplea fundamentalmente para modelar la variable longitud, mientras que una versión generalizada de ella

$$f(t) = f_{\infty} [1 - e^{-k(t-a)}]^b; \quad t \geq a; k > 0; b \geq 1, \quad (1)$$

que incluye un nuevo parámetro $b \geq 1$, permite modelar tanto la longitud como el peso de algunas especies animales en los casos de crecimiento isométrico ($b = 3$) como alométrico ($b \neq 3$).

Siguiendo la metodología desarrollada por Gutiérrez et al. (2007) en el contexto de crecimiento Gompertz, introducimos un modelo estocástico asociado con la curva generalizada de Bertalanffy que permite estudiar la evolución en el tiempo de variables de crecimiento asociadas tanto a comportamiento individual como medio de una

población. Dicho modelo es un proceso de difusión $\{X(t); t \geq t_0\}$ definido en \mathbb{R}^+ y con momentos infinitesimales

$$\begin{aligned} A_1(x, t) &= \frac{bck}{e^{kt} - c} x \\ A_2(x, t) &= \sigma^2 x^2. \end{aligned}$$

(con $c = e^{ka}$ y $\sigma > 0$) que verifica que su función media (condicionada al valor inicial) es una curva del tipo (1).

En este trabajo consideramos un proceso estocástico de dicho tipo para modelar datos de crecimiento de pez espada (*Xiphias gladius* Linnaeus, 1758) basados en un estudio realizado por J. Chong and M. Aguayo (2009) sobre algunos aspectos relativos a la edad y el crecimiento de pez espada en el Pacífico suroriental.

1. Modelo: ajuste, predicción y tiempos de primer paso

A partir de los datos de peso medio en edades comprendidas entre 1 y 17 años, para ambos sexos (ver Figura 1), se han obtenido las estimaciones máximo verosímiles de los parámetros del modelo (ver cuadro 1).

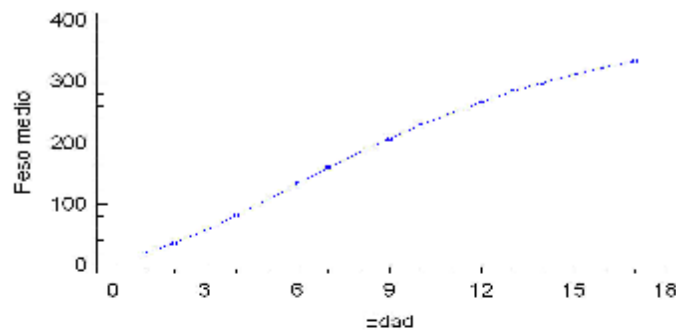


Figura 1: Peso medio relativo a la edad para pez espada en ambos sexos.

Dado que el proceso considerado verifica que $E[X(t)|X(t_0) = x_0]$ es una curva de Bertalanffy generalizada, hemos realizado el ajuste de los datos mediante la función $E[X(t)|X(t_1) = x_1]$, donde t_1 es el primer instante de observación y x_1 el peso medio observado en dicho instante. Además, y con vistas a obtener un mejor ajuste se han considerado los valores $E[X(t_i)|X(t_{i-1}) = x_{i-1}]$, $i = 1, \dots, 17$, donde x_{i-1} es el valor

| Parámetro | Valor estimado |
|-----------|----------------|
| b | 3.0631 |
| c | 0.7305 |
| k | 0.1451 |
| σ | 0.0003 |

Cuadro 1: Estimación de los parámetros del modelo usando todos los datos.

observado en t_{i-1} . La figura 2 muestra los valores observados, la estimación de la función $E[X(t)|X(t_1) = x_1]$ y de los valores $E[X(t_i)|X(t_{i-1}) = x_{i-1}]$ (unidos por líneas para una mejor visualización), $i = 1, \dots, 17$.

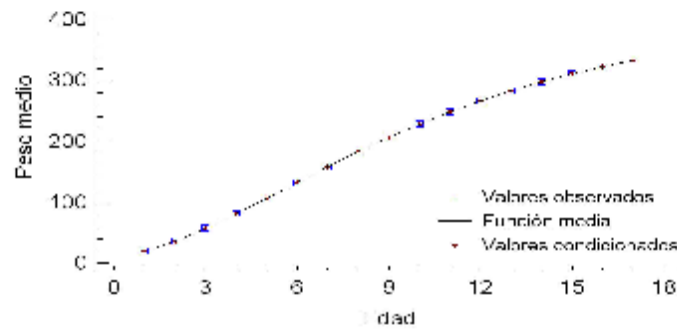


Figura 2: Valores observados, estimación de la función tendencia y de los valores $E[X(t_i)|X(t_{i-1}) = x_{i-1}]$, $i = 1, \dots, 17$.

También, para ilustrar las capacidades predictivas del modelo, se ha realizado su estimación máximo verosímil usando los datos entre los años 1 al 16 y se han realizado predicciones para el año 17. La estimación de los parámetros se muestra en el cuadro 2, mientras que en el cuadro 3 aparecen los valores observados para el año 17, predicciones puntuales dadas por la estimación de la media y la moda de la variable $X(17)|X(16) = x_{16}$, así como una predicción por intervalo proporcionada por la estimación de los cuantiles de órdenes 0.025 y 0.975 de la anterior variable.

Por otra parte, una de las ventajas de usar modelos estocásticos, como el aquí usado, para modelar variables de crecimiento es la posibilidad de estudiar algunas cuestiones relativas a su evolución en el tiempo como el logro de un cierto valor en su crecimiento o un cambio en la pauta de comportamiento. Por ejemplo, el instante en el que un animal alcanza el tamaño mínimo para su consumo, o el tiempo en el que una población alcanza un cierto tamaño, pueden ser de interés y ambos pueden formularse

| Parámetro | Valor estimado |
|-----------|----------------|
| b | 3.0589 |
| c | 0.7310 |
| k | 0.1450 |
| σ | 0.0003 |

Cuadro 2: Estimación de los parámetros del modelo sin el último dato.

| Valor observado | Media condicionada estimada | Moda condicionada estimada | Cuantiles condicionados estimados |
|-----------------|-----------------------------|----------------------------|-----------------------------------|
| 333.2 | 333.331 | 333.445 | (333.092,333.571) |

Cuadro 3: Valores observados y predichos para el peso medio de pez espada en el año 17.

como un problema de tiempo de primer paso a través de una barrera constante S , es decir, como

$$T_{S,x_1} = \text{Inf}_{t \geq t_1} \{t : X(t) > S | X(t_1) = x_1\}$$

cuya solución pasa por la obtención de la función de densidad de tal variable. En este sentido, para ilustrar esta cuestión en esta aplicación, se ha estudiado el problema del tiempo de primer paso por dos barreras, $S = 300$ y $S = 350$. En el primer caso el valor S se alcanza por los peces en el rango de tiempo considerado, mientras que el segundo se alcanza fuera de él.

La función de densidad del tiempo de primer paso es solución de una ecuación integral de Volterra de segunda especie que debe ser resuelta por procedimientos numéricos dado que no se dispone de una solución explícita en este caso (ver Gutiérrez et al. (1997)). Para ello hemos usado la metodología desarrollada en Román et al. (2008). La figura 3 muestra la función de densidad del tiempo de primer paso para los dos casos considerados, cuyos valores medios son 14.138 y 18.903, respectivamente.

Agradecimientos

Este trabajo ha sido financiado en parte por el Ministerio de Educación y Ciencia, España, bajo los proyectos MTM2008-05785/MTM y HI2007-0034, y por la Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía), España, bajo el proyecto P06-FQM-02271.

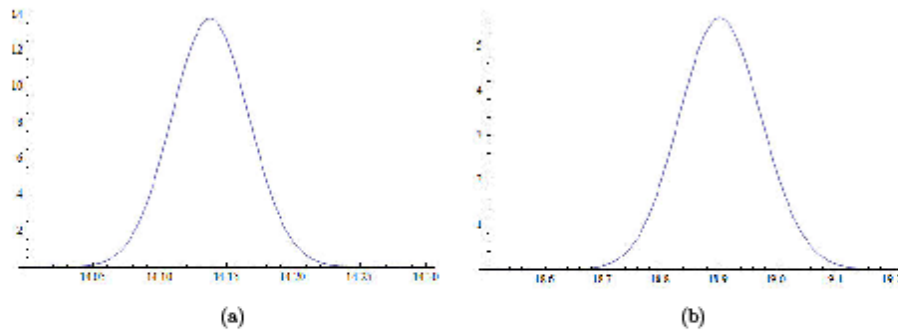


Figura 3: Densidades del tiempo de primer paso para el peso medio de pez espada a través de una barrera constante: (a) $S = 300$; (b) $S = 350$.

Referencias

- [1] Chong, J., Aguayo, M. (2009). Edad y crecimiento del pez espada (*Xiphias gladius* Linnaeus, 1758) en el Pacífico suroriental (Diciembre 1994-Septiembre 1996). *Latin American Journal of Aquatic Research* 37(1), 1-15.
- [2] Gutiérrez, R., Ricciardi, L., Román, P., Torres, F. (1997). First-passage-time densities for time-non-homogeneous diffusion processes. *Journal of Applied Probability*, 34(3), 623-631.
- [3] Gutiérrez, R., Román, P., Romero, D., Serrano, J.J., Torres, F. (2007). A new gompertz-type diffusion process with application to random growth. *Mathematical Biosciences* 208, 147-165.
- [4] Román, P., Serrano, J. J., Torres, F. (2008). First-passage-time location function: Application to determine first-passage-time densities in diffusion processes. *Computational Statistics and Data Analysis* 52, 4132-4146.

Uso de Diversas Técnicas Estadísticas en el Diseño de una Bebida con Actividad Funcional

Evangelina Ruiz-Figueroa¹, Elia Nora Aquino-Bolaños¹, Rosa Isela Guzmán-Jiménez¹, Diana Elizabeth Leyva-Daniel², Takuo Nakano³, Eryck R. Silva-Hernández¹

esilva@uv.mx

¹Instituto de Ciencias Básicas. UV

²Universidad Tecnológica de la Mixteca. Huajuapán de León, Oaxaca. México

³Departamento de Agricultura. Alimentos y Ciencias de la Nutrición. Universidad de Alberta, Canadá

Resumen

El diseño de una bebida nueva conlleva una serie de pruebas en las que los métodos estadísticos juegan un papel muy importante. Este estudio tuvo la finalidad de desarrollar una bebida para el control de peso utilizando compuestos bioactivos de plantas y animales. Los análisis estadísticos desarrollados incluyeron técnicas exploratorias básicas, análisis de correlación simple, análisis de varianza y la técnica multivariada de las caritas de Chernoff. Los resultados indicaron correlaciones negativas significativas ($p \leq 0.05$) entre el tiempo de almacenamiento y los valores de antocianinas tanto monoméricas como poliméricas. La relación negativa entre el tiempo de almacenamiento y el contenido de antocianinas poliméricas no es común, por lo que podría indicar que existe una influencia de alguno o algunos de los ingredientes sobre estas antocianinas. El análisis sensorial se realizó con 80 panelistas en base a una escala hedónica; el análisis de varianza reveló que no existen diferencias significativas ($p > 0.05$) en el nivel de agrado de la bebida entre géneros, edades, grado de estudios ni frecuencia de consumo. Sin embargo, sí se encontraron diferencias significativas ($p \leq 0.05$) para el nivel de agrado entre la posibilidad de compra del producto; los consumidores que dijeron que “definitivamente” sí comprarían la bebida mostraron un nivel de agrado promedio de 6.3 (“me gusta ligeramente”), mientras que los que dijeron “definitivamente no”, presentaron un promedio en el nivel de agrado de 4 (“me disgusta ligeramente”) y 5.37 (“ni me gusta ni me disgusta”) para los panelistas que opinaron que “no lo sabían y que tendrían que preguntar”. La técnica de las caritas de Chernoff ayudó a identificar a las personas que prefirieron o no la bebida.

Introducción

Las bebidas son una forma atractiva de incorporar componentes que promueven un beneficio a la salud, estas pueden contener compuestos bioactivos de plantas y animales tales como la Jamaica rosa (*Hibiscus sabdariffa*). A esta flor se le atribuyen diversas propiedades benéficas para la salud debido a los compuestos que contiene como el ácido protocatehúico, al cual se le atribuye una actividad diurética y, el ácido hidroxicitrico, el cual actúa al inhibir a la enzima ATP-Citratoliasa evitando así la síntesis de ácidos grasos y la formación de triacilglicéridos en la mitocondria. Además, la Jamaica rosa contiene antocianinas, las cuales son reconocidas por poseer actividad antioxidante. Por otro lado, el quitosano es un derivado del exoesqueleto de algunos crustáceos que se ha incorporado a la dieta como suplemento dietético, puesto que posee la capacidad de disminuir

marcadamente la absorción de lípidos en el tracto digestivo. La combinación del extracto de *H. sabdariffa* y quitosano podría ayudar a controlar el peso de personas que lo requirieran. De esta manera, el presente estudio tuvo como objetivo el uso de diversas técnicas estadísticas en el diseño de una bebida con actividad funcional.

Material y Métodos

Para la elaboración de la bebida se realizó un extracto acuoso de jamaica (*Hibiscus sabdariffa*) el cual se elaboró a partir de 30g de la flor y fue sometido a ebullición durante 10 minutos. Se realizó la mezcla de los siguientes ingredientes: agua, ácido cítrico, quitosano, extracto de jamaica y edulcorante. Las determinaciones realizadas fueron antocianinas monoméricas, (%) color polimérico, densidad de color, acidez titulable expresada g/L cde Ác. hidroxícitrico y °Brix. Las muestras fueron almacenadas a 4°C y analizadas cada semana por 30 días. Al inicio del almacenamiento, la bebida fue evaluada sensorialmente por 80 panelistas no entrenados mediante una escala hedónica de 9 puntos.

Con los resultados de los parámetros de estabilidad de la bebida, se construyeron bases de datos en el paquete estadístico Statistica 8.0 calculándose, medias y desviaciones estándar, análisis de varianza (ANOVA), análisis de correlación simple, además de construir un gráfico de caritas de Chernoff con los datos del análisis sensorial.

Resultados

El Análisis de Varianza y la prueba de Tukey (Tabla 1) revelaron diferencias significativas a través de las semanas en todos los parámetros estudiados con excepción de la acidez y el pH.

Tabla 1. Resultados de la evaluación de los parámetros de estabilidad de una bebida a base de Jamaica rosa y quitosano.

| Semanas | Antocianinas Monoméricas | Antocianinas Poliméricas | Densidad de Color | Acidez g/L Ác. hidroxícitrico | °Brix | pH |
|-------------|--------------------------|--------------------------|-------------------|-------------------------------|-------------|-----------|
| Tiempo Cero | 71.7±2.36a | 26.7±6.4ab | 0.94±0.01b | 0.21±0.015a | 1.08±0.0a | 3.7±0.03a |
| Semana 1 | 33.9±0.64b | 30.9±5.6a | 0.79±0.03b | 0.2±0.015 a | 1.03±0.02ab | 3.6±0.01a |
| Semana 2 | 30.5±1.1bc | 22.4±4.8ab | 2.11±0.36a | 0.21±0.015a | 1.04±0.00ab | 3.7±0.00a |
| Semana 3 | 32.4±0.84bc | 25.9±0.5ab | 0.59±0.02bc | 0.21±0.015a | 1.08±0.02a | 3.7±0.03a |
| Semana 4 | 29.1±1.0c | 16.8±0.48b | 0.29±0.01 | 0.2±0.015a | 0.97±0.05b | 3.7±0.04a |

Letras diferentes en la misma columna indican diferencias significativas a $p \leq 0.05$.

El efecto del almacenamiento de la bebida en el contenido de antocianinas monoméricas y color polimérico se muestra en la Figura 1. Ambos factores sufren una disminución de sus valores a lo largo del almacenamiento. Es importante señalar que el color polimérico tiene un comportamiento inusual ya que Hager et al. (2008) observaron que el porcentaje de color polimérico en jugos clarificados y no clarificados se ve incrementado de 12.3 a 38.6 % durante su almacenamiento y sin observarse diferencias significativas entre ambos tipos de jugo. Lo anterior podría indicar una influencia de alguno de los ingredientes sobre el color polimérico.

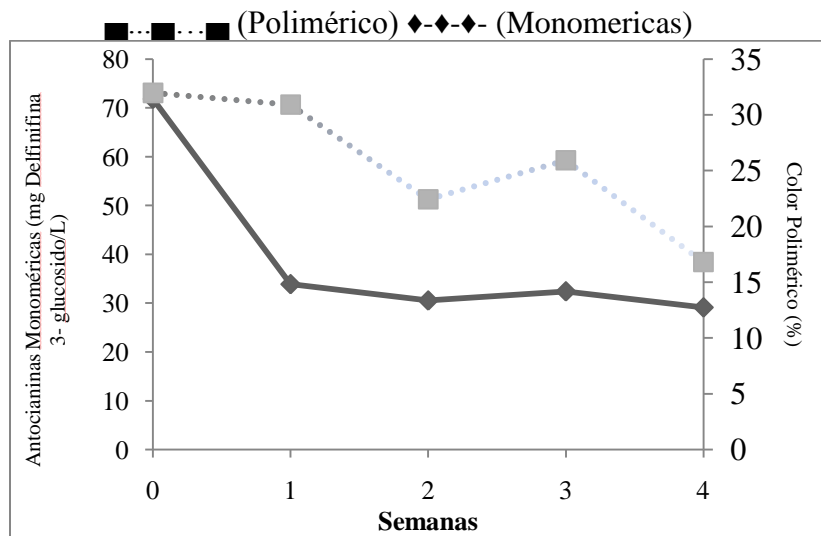


Figura 1. Antocianinas totales y porcentaje de color polimérico evaluados semanalmente durante 4 semanas en una bebida de jamaica y quitosano.

El análisis de correlación reveló que existen asociaciones negativas significativas ($p \leq 0.05$) entre el tiempo de almacenamiento y los valores de antocianinas tanto monoméricas como poliméricas. La variación observada en el porcentaje de color polimérico puede ser atribuida a los componentes que constituyen la bebida.

En el análisis sensorial, el Análisis de Varianza reveló que no existen diferencias significativas ($p > 0.05$) en el nivel de agrado de la bebida entre géneros, edades, grado de estudios ni frecuencia de consumo. Sin embargo, sí se encontraron diferencias significativas ($p \leq 0.05$) para el nivel de agrado entre la posibilidad de compra del producto; esto seguramente debido a que, aunque no les fuera agradable la bebida, el efecto benéfico a la salud podría tal vez motivarlos a comprar el producto. Los consumidores que dijeron que definitivamente sí comprarían la bebida mostraron un nivel de agrado promedio de 6.3 (“me gusta ligeramente”), mientras que los que dijeron que “definitivamente no” presentaron un promedio en el nivel de agrado de 4 (“me disgusta ligeramente”) y de 5.37 (“ni me gusta ni me disgusta”) para los panelistas que opinaron que no lo sabían que tendrían que pensarlo (Figura 2). El 75% de los panelistas dijo que seguramente sí compraría el producto (Figura 3). Lo anterior pone de manifiesto la importancia de que un producto de este tipo tenga buen sabor para su aceptación, así como la influencia de un beneficio extra sobre la probabilidad de compra.

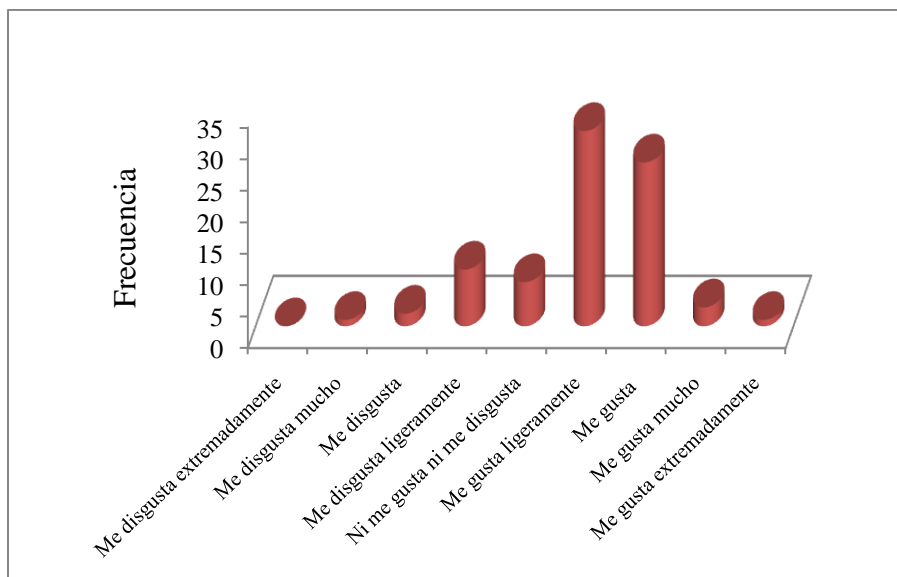


Figura 2. Resultados de la prueba sensorial afectiva de una bebida a base de Jamaica rosa y quitosano en 80 panelistas no entrenados.

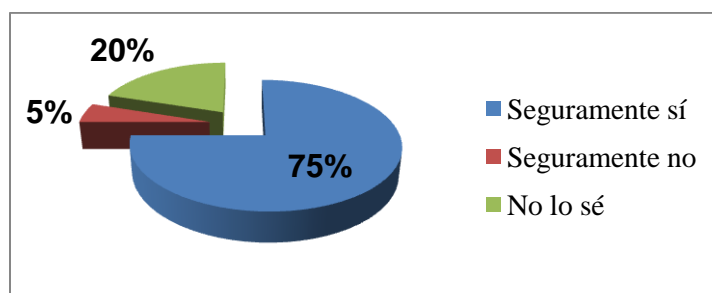


Figura 3. Resultados de la probabilidad de compra de una bebida a base de Jamaica rosa y quitosano después de informar a los panelistas sobre su funcionalidad.

La evaluación sensorial hedónica de cada panelista participante fue graficada en una carita de Chernoff (Figura 4) asignando los valores a la curvatura de la boca. La escala de 9 puntos va de 1=me disgusta extremadamente a 9=me gusta extremadamente. En la figura 4 se aprecia que los dos primeros renglones muestran caras negativas (desagrado a la bebida), mientras que el resto son positivas (agrado por la bebida). Lo anterior indica una aceptación de la bebida de más del 60% de los participantes.

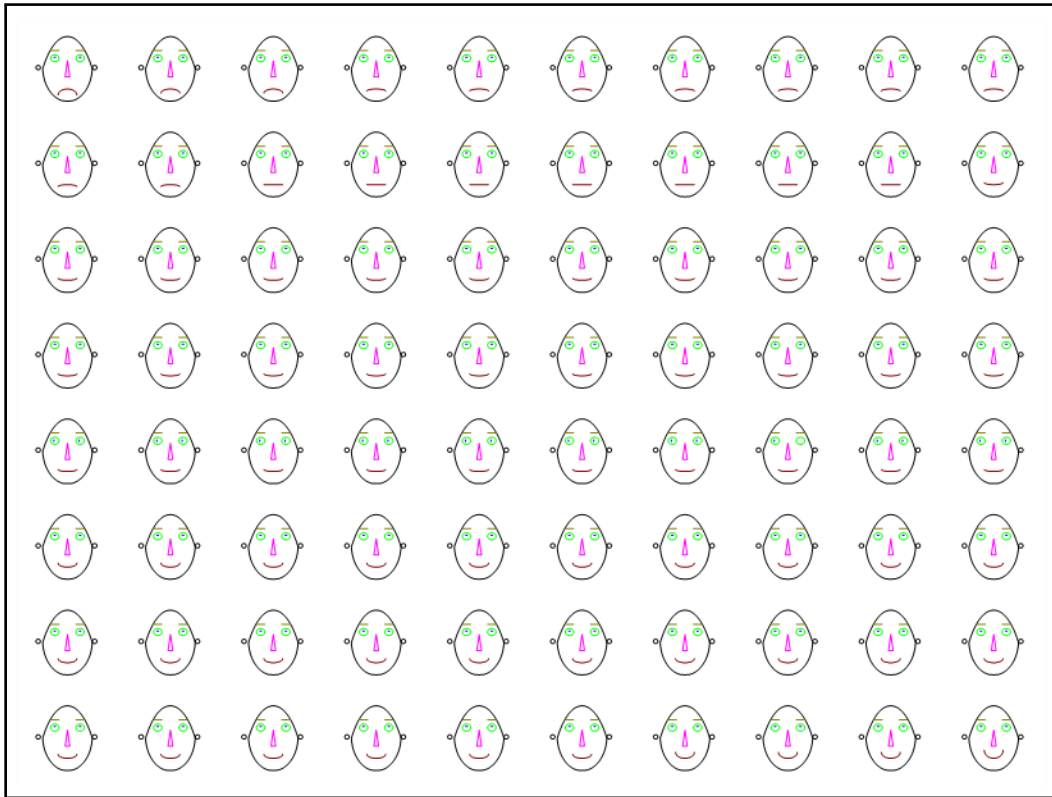


Figura 4. Caritas de Chernoff de los 80 panelistas participantes. La escala hedónica de 9 puntos está asignada a la curvatura de la boca.

Conclusión

El uso de diversas técnicas estadísticas en el diseño de productos alimenticios como la bebida funcional de este estudio, es fundamental para explorar, comparar e inferir sobre los datos obtenidos. Del mismo modo, el empleo de los métodos estadísticos es un gran auxiliar para tomar la mejor decisión tanto en el estudio presente, como en la proyección hacia estudios futuros similares.

REFERENCIAS

- Anzaldúa- Morales A. (1994). *La evaluación sensorial de los alimentos en la teoría y en la práctica*. España: Acribia.
- Giustin, M.M. and Wrolstad, R. (2001). *Characterization and Measurement With UV-Visible Spectroscopy*. Protocols in Food Analytical Chemistry, Wrolstad, R. E. Anthocyanins.
- Hager, J. C., Howard, R., and Prior L. R. (2008). Processing and Storage Effects on Monomeric Anthocyanins, Percent Polymeric Color, and Antioxidant Capacity of Processed Blackberry Products. *Agriculture and Food Chemistry* **56**, 689-695.
- Pedrero, D y Pangborn, R. M. (1997). *Evaluación sensorial de los alimentos*. México: Alambra Mexicana, 103- 107.

Aplicación del Algoritmo Chaid para el Analisis de la Televisión y el Comportamiento en Niños Preescolares

Laura Oliva Zárate¹, Claudio Castro²

¹ Académico del instituto de Psicología y Educación. UV. loliva@uv.mx

² Académico de la Facultad de Estadística e Informática. UV. ccastro@uv.mx

Resumen

Se presenta un estudio que realiza un análisis del uso de la televisión en niños de 4 a 5 años de edad y su relación con los problemas de conducta. Los instrumentos utilizados fueron el Cuestionario sobre el comportamiento de niños de 1½-5 años de T. Achenbach, T. y Rescorla, (2002), aplicado a 535 padres de niños de 20 escuelas de nivel preescolar que resultaron de un muestreo representativo de la ciudad de Xalapa-México a fin de identificar problemas conductuales en los infantes, el Cuestionario para padres sobre la televisión y su relación sobre los problemas de conducta, la Entrevista sobre preferencias televisivas en el niño preescolar y el Grupo de Discusión con Maestros de Educación Preescolar. El análisis realizado es tanto descriptivo como multivariante, especialmente obteniéndose resultados concluyentes con la segmentación CHAID, la cual demuestra una relación directa entre el tiempo de visión y los programas televisivos respecto a los problemas de conducta en los infantes, marcándose la influencia en el entorno familiar y el estrato socioeconómico.

Introducción

Entre los diversos estudios realizados respecto las características predominantes de los infantes televidentes encontramos como variables de estudio la edad de inicio, el tiempo de visión, los cambios en el tiempo de visión respecto a la edad, el género, las preferencias televisivas, motivación para ver la televisión, los cuales han sido constantemente estudiadas, como lo apunta Singer (2001).

Existe una enorme cantidad de estudios sobre la influencia de la televisión en los niños, la gran mayoría producidos en Estados Unidos. “Quizá el hallazgo más impresionante –apunta Roberts (En Sánchez, 1989, p.28)– en la investigación sobre los niños y los medios es la enorme cantidad de aprendizaje que parece tener lugar, aun a edades tempranas”. Esta es ya una constante en la investigación sobre la televisión y “la conclusión general es que los espectadores aprenden de la televisión y que este aprendizaje influye en su conducta” (Fernández, C., Baptista, P. y Elkes, D. (1991).

En México son escasos los estudios realizados específicamente sobre la influencia de la televisión en las nuevas generaciones; algunos estudios son de Fernández y cols., Corona y Cornejo; y la recopilación que hicieron García Silberman y Ramos (En Martínez, 2002, p.73), los cuales sientan un precedente que requiere ser continuado por la innegable influencia de la televisión sobre el niño.

Por otra parte, en lo que se refiere a las causas que originan los problemas conductuales hasta el momento no han sido adecuadamente determinadas. Aunque algunas pueden incluir factores tales como la herencia, fisiológicos, experiencias de vida, situaciones

traumáticas y el funcionamiento familiar ningún estudio ha podido demostrar que alguno de estos factores sea la causa directa de los problemas del comportamiento. En lo que respecta a los efectos de la televisión en niños, la mayoría de los estudios realizados y su relación con los problemas infantiles se ha explorado ampliamente con la agresividad, hiperactividad y la obesidad, explorando escasamente otras conductas.

Entre las investigaciones que mencionan a las alteraciones de conducta, encontramos un estudio realizado por Medrano y colaboradores (1992) sobre los efectos de la televisión en niños de 5 a 14 años de edad. Encontrándose que la televisión, tuvo una notable influencia en la conducta de los niños en edad escolar en dicho estudio. Un 50% de los niños en edad escolar en un barrio de clase socioeconómica baja, que veían diariamente TV, exhibían alteraciones en su conducta, prefiriendo los varones (73.8%) los programas de contenido violento. Por otra parte, en un barrio de clase media donde el 48% de los niños en edad escolar ven TV 4 horas diarias, el 80% ingería alimentos mientras observaba la TV, lo cual sin duda contribuye a una temprana obesidad en ellos.

Hasta ahora, teorías como la del aprendizaje social (Bandura, 1977) se han avocado principalmente a la conducta agresiva, siendo ésta de las más estudiadas en donde se explica cómo las personas adquieren las pautas básicas sobre las consecuencias de las conductas a través de la observación de modelos. No obstante, existen otras corrientes que interponen rasgos de carácter individual, social o cultural que condicionan, y por tanto relativizan, la recepción individual. Lo cierto es que desde hace varias décadas algunos problemas conductuales han sido catalogados como influenciados por la televisión tal es el caso de la hiperactividad y los problemas del sueño. Como podemos apreciar existen corrientes con distintas posiciones respecto a lo que sería la influencia de la televisión, lo cual es importante integrar a fin de comprender este fenómeno y muy en particular el tema que nos ocupa respecto a su relación con los problemas conductuales.

Es importante aclarar, que es probable que no todos los problemas conductuales y emocionales sean influenciados por este medio, definitivamente el comportamiento humano no depende sólo de la imitación, pero entonces... qué otros mecanismos podrían atribuirse a la televisión para relacionarla con algún problema de conducta si este fuera el caso. De ahí la relevancia del presente estudio, el cual pretende identificar cuáles comportamientos serían más influenciados que otros por este medio, ya que si bien se han reportado cientos de trabajos de investigación que revelan la existencia de relaciones entre determinadas conductas con la televisión, algunos no han sido exploradas en su totalidad. Y es que de antemano sabemos la dificultad investigar la influencia televisiva aislada, al margen del resto de factores contextuales en que se inserta la vida diaria de cualquier niña o niño: entorno familiar, hábitat, estrato y escuela, entre otros. Por ello, resulta de gran relevancia aplicar para este estudio modelos estadísticos que nos permitan acercarnos a la explicación de los problemas conductuales con respecto al uso de la televisión a fin de dilucidar las preguntas que aún están sin contestar.

Metodología

La muestra estuvo conformada por 535 niños de 4 y 5 años (52% niñas y 48% niños) de 20 escuelas de nivel preescolar que resultaron de un muestreo representativo de la ciudad de

Xalapa-México, considerándose tanto escuelas públicas como privadas, así como los diferentes estratos socioeconómicos.

Los instrumentos utilizados fueron el Cuestionario sobre el comportamiento de niños de 1½-5 años de T. Achenbach, T. y Rescorla, (2002) a fin de identificar problemas conductuales en los infantes y el Cuestionario para padres sobre la televisión y su relación sobre los problemas de conducta. Dichos instrumentos fueron contestados por los padres de los niños de la muestra dentro de las instalaciones escolares.

Si bien la utilización de estadísticos univariados y bivariados se hacían pertinentes para el análisis a fin de obtener frecuencias y relaciones entre dos variables, resultaba necesario avocarnos a un análisis de mayor profundidad a fin de alcanzar los objetivos propuestos, por lo que se optó por la aplicación del algoritmo de segmentación CHAID (*Chi-square Automatic Interaction Detection*), propuesto por Kass (1980). Lo anterior, debido a que se presenta el problema de la descripción parsimoniosa de un conjunto numeroso de datos que constituyen la caracterización de un gran número de individuos, según un gran número de variables.

Y es que el análisis de segmentación (AS) reduce la complejidad del problema, rechazando tabulaciones cruzadas no significativas, detectando automáticamente los mejores predictores y creando subgrupos potencialmente explicativos de la variable dependiente. Los beneficios de segmentar una población o muestra son muchos, entre los cuales; construir un perfil más preciso de los individuos, que componen un colectivo bajo estudio, agrupar para conocer mejor un subgrupo poblacional y obtener mejores pronósticos, sobre el comportamiento de grandes grupos de datos.

Por tanto, los métodos de segmentación, son una de las áreas que compete atender a la Estadística, ya que el término segmentar significa dividir una población en segmentos homogéneos, en relación a una o varias características (que no son otra cosa que variables estadísticas). El AS es una técnica que a partir de la información que suministran, ciertas variables independientes o explicativas, clasifica un conjunto de objetos en grupos, capaces de describir de la mejor manera posible la variable dependiente.

El Algoritmo de Segmentación CHAID, es considerado un algoritmo general de segmentación, se utiliza con fines exploratorios y descriptivos, con el objetivo fundamental de encontrar la partición de una muestra de objetos en grupos, capaces de describir de la mejor manera posible la variable dependiente. Supone que la variable dependiente es categórica y utiliza la prueba Chi-cuadrado, para contrastar independencia, en distintas fases del proceso.

CHAID segmenta la muestra en dos o más grupos, en donde se busca que los objetos que pertenecen al mismo grupo, sean lo más homogéneo posible y objetos de diferente grupo son heterogéneos. Tomando como base las categorías del mejor predictor (variable independiente), de una variable dependiente, divide cada uno de estos grupos en subgrupos más pequeños, en base a otras variables independientes. Este proceso iterativo de partición continúa, hasta no encontrar ninguna variable independiente estadísticamente significativa. CHAID muestra los segmentos finales, en forma de diagrama de árbol. Los segmentos que

CHAID construye, son mutuamente exclusivos y exhaustivos, es decir los segmentos no se sobreponen, y cada objeto de la muestra está contenida exactamente en un segmento.

Entre los resultados obtenidos más relevantes en el estudio podemos señalar que los 247 niños detectados con algún problema conductual pertenecen en su mayoría (51.4%) al estrato bajo, seguido por el nivel medio con un 35.6% y finalmente con un 13% al estrato alto. Respecto al sexo notamos mínimas diferencias al igual que en los que no presentaron algún problema, obteniéndose un 51.4% para las niñas y un 48.6% para el sexo masculino. Finalmente respecto a la edad resultaron los niños de 4 años con un 38.9% mientras que aparecieron con 60.7% los infantes de 5 años, una cifra mayor que la de los niños de 4.

Respecto al uso de la televisión el más alto porcentaje (57%), revela que dos y tres horas al día es el tiempo que la mayoría de los niños ven tv., y con la aplicación de la prueba chi-cuadrada no se encontraron diferencias en cuanto a estrato, edad y sexo. El 56.9% de los padres acepta que los niños ven con ellos telenovelas y noticieros y que la mayoría de los niños prefieren hacer otra cosa que ver la televisión. Quien elige el programa a ver es en la mayoría el mismo niño y la mayoría de los padres no controla el tiempo de visión de sus hijos y no han tomado medidas respecto a su uso. Y el 68% de los padres expresa la existencia de alcoholismo o drogadicción en la familia.

Por medio del estadístico CHAID se obtiene que en las relaciones familiares malas, el 97% de los niños manifiestan problemas conductuales. Y aunque las Relaciones familiares sean buenas, pero sin reglas claras en casa, el 70% de los niños manifiestan agresividad o problemas de atención. No encontrándose una relación directa entre el tiempo de visión y los programas televisivos respecto a los problemas de conducta en los infantes, marcándose la influencia en el entorno familiar y estratosocioeconómico.

Las conclusiones nos remiten a encontrar relación entre los problemas conductuales y el entorno familiar principalmente así como con el estrato, y respecto a la televisión sólo a la imitación de ciertas conductas que no necesariamente implican la variedad de conductas que involucran a un problema conductual sino a comportamientos aislados.

De esta manera se observa la complejidad del estudio del comportamiento humano, ya que los análisis estadísticos aplicados particularmente con el CHAID nos explican la conducta de maneras insospechadas que incluso en ocasiones no son claras o parecieran irrelevantes, mas sin embargo nos provocan la reflexión y el cambio de rutas para explicarnos la naturaleza humana. Aquí por ejemplo el entorno familiar continúa siendo el de mayor peso con respecto a los problemas de conducta, lo que nos llevaría a replantearnos la influencia de la televisión.

REFERENCIAS

- Achenbach, T. y L. Rescorla (2002). *Cuestionario sobre el comportamiento de niños (as) de 11/2-5 años*. Burlington: ASEBA.
- Bandura, A. (1977). *Teoría Social del aprendizaje social*. Madrid: Espasa- Calpe.
- Fernández, C., Baptista, P. y Elkes, D. (1991). *La televisión y el niño*. México Nueva Biblioteca Pedagógica.

- Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 119-127.
- Martinez, I. (2002). *¿Quién decide lo que ven los niños?* México: Pax.
- Medrano, A.M., Lockward, J.G., Jiménez, M. y Alcántara Y. (1992). ***Efectos de la televisión en niños. Acta Médica Dominicana 20 (5), 23-30.***
- Sánchez, E. (1989). *Teledicción infantil: ¿mito o realidad?* México: Universidad de Guadalajara.
- Singer, D.G. (2001). *Handbook of Children and the Media*. U.S.A: Sage Publications.

Comparando el Análisis de Componentes Principales (PCA), el Test No Paramétrico de Friedman-Nemenyi (TFN) y el Análisis Discriminante (AD) para la Clasificación de Expresiones Faciales en Bebés Cuando Toman una Fórmula Infantil

Rosa Guadalupe Herrera-Lee^{1,2}, Iñigo Verdalt-Guzmán³, Takuo Nakano⁴, Lech Ozimek⁴, María Luz Márquez-Barradas⁵, Homero Vladimir Ríos-Figueroa⁶, Jesús Ofelia Angulo-Guerrero¹, Eryck R. Silva-Hernández³

¹ Instituto Tecnológico de Veracruz. México. roherrera@gmail.com

² Facultad de Nutrición-Xalapa. UV. México

³ Instituto de Ciencias Básicas. UV. México

⁴ Departamento de Agricultura. Alimentos y Ciencias de la Nutrición. Universidad de Alberta, Canadá

⁵ Instituto de Investigaciones Psicológicas. UV. México

⁶ Departamento de Inteligencia Artificial. UV. México

Resumen

La evaluación sensorial de alimentos en adultos es relativamente sencilla, ya que la explicación del procedimiento a seguir y la retroalimentación de los evaluadores se desarrolla mediante una comunicación verbal y/o escrita. Sin embargo, cuando se desea realizar una evaluación sensorial en bebés muy pequeños, la comunicación no puede ser ni verbal ni escrita, por lo que el uso de otras técnicas, como la evaluación de las expresiones faciales, es necesario para obtener información. Las expresiones faciales de bebés se han empleado como una herramienta para determinar la valencia (positiva o negativa) y la intensidad del confort, aunque existe poca información disponible sobre estas escalas de agrado-desagrado (hedónicas) y cómo se han validado estadísticamente. De esta manera, el presente estudio compara tres técnicas estadísticas (PCA, TFN y AD) para la construcción de una escala hedónica basada en las expresiones faciales de bebés entre 4 y 6 meses de edad al ingerir diferentes fórmulas infantiles. Se reclutaron a 30 bebés en un Hospital de la Ciudad de Xalapa, México. A cada bebé se le proporcionaron aleatoriamente dos tipos de fórmulas infantiles, una a base de leche de vaca y otra a base de soya. Cada toma fue fotografiada y videofilmada. Se seleccionaron las fotografías más representativas y claras de cada una de las 60 tomas, quedando sólo 22 fotografías, todas de bebés diferentes. El formato a color fue transformado en monocromático para reducir el sesgo. El análisis de las fotografías se realizó en dos etapas. En la primera, se les pidió a 37 mujeres de entre 16 y 40 años de edad, que formaran 3 grupos con las fotografías: a) fotografías expresando agrado por la fórmula (expresión positiva), b) expresando desagrado (expresión negativa) y c) indiferencia (expresión neutra). Los resultados fueron analizados por ANOVA y Tukey, además de un PCA, determinándose que 7 de las fotografías representaban una escala hedónica. Para la segunda etapa, las 7 fotografías se mostraron a 60 mujeres entre 16 y 40 años, a las cuales se les pidió que las ordenaran de mayor a menor agrado. Los resultados se analizaron mediante el TFN. Los 60 videos fueron analizados por 43 observadores pidiéndoles que evaluaran el tiempo que el bebé mantiene el biberón en la boca, las veces que lo rechaza y las conductas que presenta el bebé durante la toma. Los datos de los videos fueron procesados mediante un AD. El PCA y el TFN revelaron una escala hedónica de 4 puntos. Sin embargo, una fotografía clasificada como “débilmente negativa” por el PCA,

fue clasificada como “débilmente positiva” por el TFN, lo cual podría indicar la existencia de una expresión neutra. El modelo desarrollado mediante el AD clasificó correctamente a casi el 90% tanto de las expresiones positivas como negativas, pero solo al 14% de las expresiones neutras. La escala hedónica sugerida incluye expresiones fuertemente positivas, débilmente positivas, débilmente negativas y fuertemente negativas, aunque más investigaciones deberían desarrollarse en esta área para determinar si las expresiones “débiles” deberían en realidad ser clasificadas como “neutras”.

Introducción

Las expresiones faciales representan una parte muy importante de la comunicación no verbal en la vida diaria de las personas y representa un medio de comunicación más rápido que el lenguaje. Con una expresión facial se puede, por ejemplo, inferir estados de la mente, entender opiniones y actitudes de otras personas (Batty y Taylor, 2003). Sin embargo, en los bebés las expresiones faciales se convierten en el principal medio de expresión de sus emociones de agrado o desagrado. En este sentido, las expresiones faciales de bebés se vuelven una herramienta poderosa para la obtención de información, especialmente cuando se trata de una evaluación sensorial.

Las expresiones faciales de bebés se han clasificado en la intensidad de confort o agrado (expresiones positivas) y de incomodidad o desagrado (expresiones negativas), caracterizando a la sonrisa como una expresión positiva y al llanto como negativa (Messinger, 2002). Sin embargo, la habilidad de interpretación de las expresiones faciales de los bebés por los adultos puede variar, especialmente respecto al género. Las mujeres son significativamente más rápidas y consistentes en discriminar o identificar expresiones faciales en los bebés que los hombres; parece ser una habilidad específica adquirida por las mujeres por una probable predisposición biológica (Proverbio *et al.*, 2007).

Existe relativamente poca información disponible sobre escalas de agrado-desagrado (hedónicas) destinadas para su uso en bebés y cómo se han validado estadísticamente, por lo que en este estudio se propone la comparación de tres técnicas estadísticas: el Análisis de Componentes Principales, el Test no paramétrico de Friedman-Nemenyi y el Análisis Discriminante) para la construcción de una escala hedónica basada en las expresiones faciales de bebés de entre 4 y 6 meses de edad al ingerir diferentes fórmulas infantiles.

Material y Métodos

Se reclutaron a 30 bebés de entre 4 y 6 meses de edad en el Hospital Civil de la Ciudad de Xalapa, Veracruz. Todos los bebés considerados estaban clínicamente sanos. A todas y cada una de las madres se les solicitó autorización para la inclusión de su hijo (a) en el estudio. A cada bebé se le proporcionaron aleatoriamente dos fórmulas infantiles, una a base de leche de vaca y otra a base de soya, limpiando la lengua entre cada toma. En total se incluyeron nueve fórmulas infantiles: cinco a base de leche y cuatro a base de soya. Cada toma fue fotografiada y video filmada.

Se seleccionaron las fotografías más representativas y claras de cada una de las 60 tomas, quedando sólo 22 fotografías, todas de bebés diferentes. El formato de las fotografías (color) fue transformado en monocromático (blanco y negro) para reducir el sesgo debido al color. El análisis de las fotografías se realizó en dos etapas. En la primera, se les pidió a

37 mujeres de entre 16 y 40 años de edad que tuvieran bebés o que convivieran con bebés al menos ocho horas al día, que formaran 3 grupos con las fotografías de acuerdo a lo siguiente: a) fotografías expresando agrado por la fórmula (expresión positiva), b) expresando desagrado (expresión negativa) y c) indiferencia (expresión neutra). Los resultados fueron analizados en el paquete estadístico Statistica 8.0 mediante Análisis de Varianza y la prueba de Tukey, además de un Análisis de Componentes Principales, determinándose que 7 de las fotografías representaban una escala hedónica.

Para la segunda etapa, las 7 fotografías se mostraron a 60 mujeres entre 16 y 40 años que tuvieran bebés o que convivieran con bebés al menos ocho horas al día, a las cuales se les pidió que las ordenaran de mayor a menor agrado. Los resultados se analizaron por la prueba no paramétrica de Friedman y el procedimiento de comparaciones múltiples pareadas de Nemenyi empleando el paquete estadístico XLSTAT 2.02.

Los 60 videos fueron analizados por 43 observadores “ingenuos” pidiéndoles que evaluaran el tiempo que el bebé mantiene el biberón en la boca, las veces que rechaza el biberón y las conductas que presenta el bebé durante la toma. Los datos de los videos fueron procesados mediante un Análisis Discriminante en el paquete estadístico Statistica 8.0.

Resultados

La figura 1 muestra los promedios de la evaluación de las 22 fotografías seleccionadas de las 60 tomas, en la que se aprecian dos grupos marcados claramente, uno formado por fotografías que muestran bebés con expresiones positivas y otro donde se muestran bebés con expresiones negativas. El Análisis de Varianza reveló diferencias significativas ($p \leq 0.05$) entre las fotografías mostrando expresiones positivas respecto a las negativas. Es de destacarse que 3 de las fotografías fueron evaluadas, en promedio, como más cercanas a un valor de 2 (correspondiente a una expresión neutra) que a 1 (expresión positiva) o a 3 (expresión negativa). Lo anterior podría sugerir la existencia de un valor neutro.

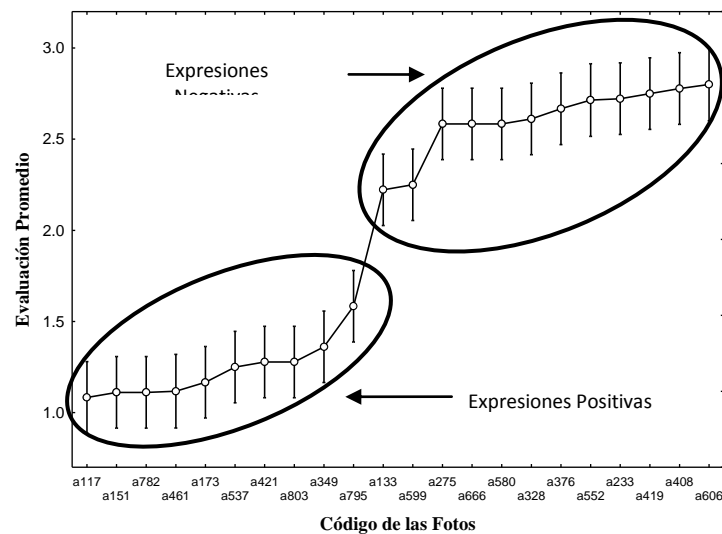


Figura 1. Promedios de la evaluación de las 22 fotografías seleccionadas de las 60 tomas. Las barras verticales indican intervalos de confianza al 95%.

El Análisis de Componentes Principales mostró al factor 1 como el de mayor influencia sobre la varianza ($\approx 64\%$), casi 10 veces mayor que el siguiente factor más importante (factor 2: 6.85%). De esta manera, se observó que las panelistas distinguieron 4 grupos de expresiones: a) fuertemente positivas (de +4 a +6 en el factor 1), b) débilmente positiva (alrededor de +2 en el factor 1), c) débilmente negativas (alrededor de -2 en el factor 1) y d) fuertemente negativas (de -4 a -6 en el factor 1), lo cual concuerda con lo sugerido por Proverbio *et al.* (2007).

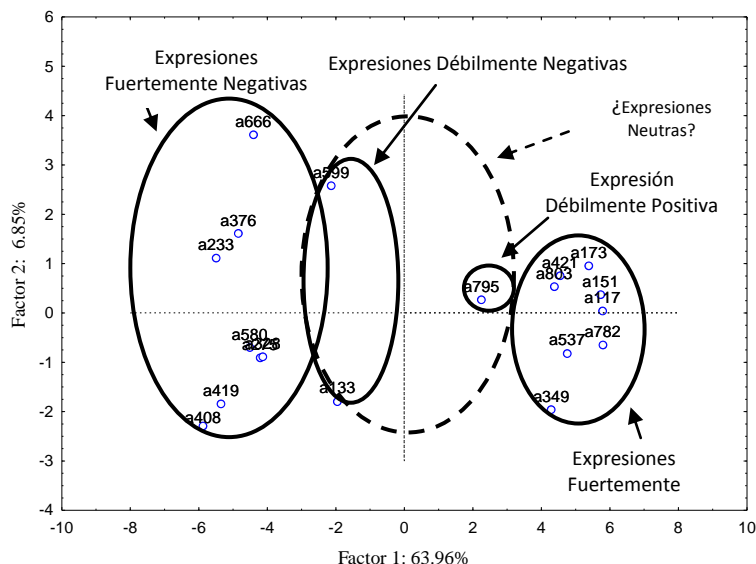


Figura 2. Gráfico de los factores 1 y 2 del Análisis de Componentes Principales de las evaluaciones de las expresiones faciales de 22 fotografías de bebés ingiriendo una fórmula infantil. Los valores junto a cada factor indican el porcentaje de la varianza total explicado por ese factor específico.

La jerarquización de las fotografías (Figura 3) analizada mediante el test de Friedman-Nemenyi confirmó la tendencia de la percepción de las panelistas al mostrar claramente los grupos de expresiones faciales “fuertemente positivas” (“a” en la figura 3) y “fuertemente negativas” (“c” en la figura 3); sin embargo, es interesante señalar que la fotografía 133 inicialmente percibida como “débilmente negativa” en el PCA, sea ahora clasificada como “débilmente positiva”.



Figura 3. Fotografías jerarquizadas. El orden va de izquierda (expresión más positiva) a derecha (expresión más negativa). Los números indican los códigos de las fotografías. Letras diferentes indican diferencia significativa a $p \leq 0.05$.

El modelo desarrollado mediante el Análisis Discriminante (AD) clasificó correctamente a casi el 90% tanto de las expresiones positivas como negativas, pero solo al 14% de las expresiones neutras. Al correr nuevamente el AD con 5 niveles (2 negativos, uno neutro y 2 positivos), el porcentaje de clasificaciones correctas disminuyó notoriamente, especialmente en la parte negativa. Se determinó que la falta de ajuste se debía a que los panelistas estaban percibiendo expresiones en una escala no balanceada, es decir, solo un

nivel negativo y 2 positivos (fuerte y débil). Al analizar de nuevo los datos, el modelo clasificó correctamente a alrededor del 80% de las expresiones tanto negativas como fuertemente positivas y cerca del 60% de las expresiones débilmente positivas.

Conclusión

Es clara la existencia de expresiones faciales tanto “positivas” como “negativas” en bebés de 4 a 6 meses de edad cuando consumen una fórmula infantil. Sin embargo, aún no está claro si todas las expresiones “débiles” deberían clasificarse como tales. Es también incierto si las técnicas estadísticas propuestas aquí son suficientemente sensibles para clasificar correctamente las respuestas de los panelistas o si en realidad lo indeterminado de la clasificación de una expresión determinada depende de los mismos panelistas en turno. Más investigaciones deben desarrollarse en esta área.

REFERENCIAS

- Batty, M., y Taylor, M.J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research* **17**, 613-620.
- Messinger, D.S. (2002). Positive and Negative: Infant Facial Expressions and Emotions. *Current Directions in Psychological Science* **11**, 1-6.
- Proverbio, A.M., Matarazzo, S., Brignone, V., Zotto, M., and Zani A. (2007). Processing valence and intensity of infant expressions: The roles of expertise and gender. *Scandinavian Journal of Psychology* **48**, 477-485.

Análisis de Parámetros Fisicoquímicos, Color, Contenido de Polifenoles, Antocianinas y Capacidad Antioxidante en Licores de Mora mediante Técnicas Multivariadas

Diana Elizabeth Leyva-Daniel, Iñigo Verdalt-Guzmán², Eryck R. Silva-Hernández², Elia Nora Aquino Bolaños², Carlos Guillermo Hernández¹, Rosa Isela Guzmán-Gerónimo²

¹ Universidad Tecnológica de la Mixteca. Oaxaca, México

² Instituto de Ciencias Básicas. UV. México

Resumen

El objetivo del presente trabajo fue evaluar la composición fisicoquímica, color, antocianinas monoméricas, color polimérico y capacidad antioxidante en licores de mora, y determinar si existen diferencias globales entre los licores mediante el análisis multivariante de la varianza (MANOVA), así como identificar asociaciones de los parámetros fisicoquímicos y color con el contenido de polifenoles, antocianinas monoméricas, color polimérico y capacidad antioxidante utilizando un análisis canónico. Se adquirieron 9 licores comerciales en la región de Xico, Veracruz y se les determinó el pH, azúcares totales, grado alcohólico y acidez, los parámetros colorimétricos L, a, b, el ángulo de matiz (h) y saturación (C), polifenoles por el método de Folin-Ciocalteu, antocianinas monoméricas y color polimérico mediante el método diferencial de pH y capacidad antioxidante por DPPH. Los datos obtenidos se analizaron utilizando el paquete estadístico Statistica 8.0. Los resultados obtenidos mediante MANOVA indicaron que, con excepción de 2 muestras, si existen diferencias globales estadísticamente significativas entre los licores evaluados. El análisis canónico indicó que a menor pH y mayor porcentaje de azúcares totales hay un mayor contenido de antocianinas monoméricas y menor color polimérico (raíz 1). De igual manera se encontró una correlación positiva entre la acidez y la densidad de color (raíz 2).

Palabras Clave: antocianinas, parámetros fisicoquímicos, capacidad antioxidante, técnicas multivariadas.

I. Introducción

Diversos estudios indican que las antocianinas, pigmentos que pertenecen a los compuestos polifenoles, poseen propiedades antiinflamatorias, antivirales, antimicrobianas y antioxidantes, además de aportar un color característico que puede considerarse como un parámetro de calidad en los alimentos. Estas sustancias bioactivas se encuentran en frutos como las zarzamoras, la cuales crecen en forma silvestre en la región de Xico, Veracruz y se utilizan para la elaboración de una bebida alcohólica denominada licor de mora. Investigaciones previas en vinos sugieren que el pH, contenido de azúcares y alcohol influyen en el contenido de las antocianinas. Datos reportados en la literatura indican que durante el añejamiento de los vinos, el color se estabiliza debido a la formación de antocianinas poliméricas. Así mismo, estudios realizados en vinos tintos jóvenes indican que casi la totalidad del color está dado por antocianinas monoméricas, pero después de un año, al menos el 50% del color percibido es debido a las antocianinas poliméricas, estos pigmentos son menos afectados por el pH, por que le dan estabilidad al color del vino

(Casassa and Catania, 2006). Por otra parte, investigaciones previas sugieren que la adición de sacarosa en fresas congeladas ayuda a mantener en mayor grado el contenido de antocianinas monoméricas y minimiza la formación de antocianinas poliméricas (Wrolstad et al., 1990). En cuanto a la actividad antioxidante, se ha reportado que durante el añejamiento de los vinos el porcentaje de antocianinas poliméricas contribuyen en gran medida a esta propiedad biológica ((Tsai et al., 2004). Si bien, numerosas investigaciones han sido realizados en vinos, la información que existe en cuanto a los licores de mora que se producen en México es escasa, así mismo no existen datos acerca de su composición fisicoquímica y su relación con el contenido de antocianinas y su efecto en la actividad antioxidante y calidad sensorial en cuanto al color. El objetivo del presente trabajo fue evaluar la composición fisicoquímica y color en licores de mora, así como el contenido antocianinas monoméricas, color polimérico y capacidad antioxidante, y determinar si existen diferencias globales entre los licores mediante el análisis multivariante de la varianza (MANOVA), así como identificar asociaciones de los parámetros fisicoquímicos y color con el contenido de polifenoles, antocianinas monoméricas, color polimérico y capacidad antioxidante utilizando un análisis canónico.

2. Materiales y Métodos

2.1 Análisis fisicoquímicos

Se adquirieron 9 licores comerciales en la región de Xico, Veracruz y se les determinó el pH, azúcares totales, grado alcohólico y acidez de acuerdo a la metodología de la AOAC (1997).

2.2 Determinación de color

Los parámetros colorimétricos L, a y b fueron evaluados utilizando un equipo Hunterlab a partir de los cuales se calculó el ángulo de matiz (h) y saturación (C) (Calvo y Durán, 1997).

2.3 Evaluación de antocianinas y capacidad antioxidante

La concentración de polifenoles totales se determinó por el método de Folin-Ciocalteu (1927), antocianinas monoméricas y el color polimérico mediante el método diferencial de pH (Giusti and Wrolstad, 2001) y la capacidad antioxidante por DPPH (Brand-Williams et al., 1995).

2.4 Análisis estadístico

La evaluación de la información obtenida se realizó mediante el paquete estadístico Statistica 8.0. En este caso, con el objetivo de determinar si los parámetros analíticos evaluados constituyen una herramienta para establecer si existen diferencias globales entre los licores, se realizó el análisis multivariado de la varianza ($p < 0.05$). Así mismo, se aplicó el análisis canónico con la finalidad de establecer asociaciones entre los parámetros fisicoquímicos y color con el contenido de polifenoles, antocianinas y actividad antioxidante.

3. Resultados

3.1 Prueba Multivariada de Wilks

La importancia del empleo de este análisis multivariado radica en que ayuda a comparar las variables evaluadas en este estudio y determinar si existen diferencias globales significativas. Mediante esta prueba se obtuvo, como se observa en la Tabla 1, los valores de significancia, p, los cuales indican que existe diferencia significativa entre ellos, es decir las características de los licores cambian dependiendo del productor, sin embargo el licor L8 no muestra diferencia significativa con respecto a la muestra L4. Lo anterior puede atribuirse a que el proceso de elaboración de los licores varía dependiendo del productor, por lo que encontramos diferentes cada uno de ellos.

| Licores | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|----|
| L1 | | | | | | | | | |
| L2 | 0.0025 | | | | | | | | |
| L3 | 0.0018 | 0.0010 | | | | | | | |
| L4 | 0.0001 | 0.0006 | 0.0070 | | | | | | |
| L5 | 0.0198 | 0.0002 | 0.0017 | 0.0010 | | | | | |
| L6 | 0.0025 | 0.0011 | 0.0011 | 0.0012 | 0.0083 | | | | |
| L7 | 0.0230 | 0.0027 | 0.0015 | 0.0004 | 0.0008 | 0.0067 | | | |
| L8 | 0.0098 | 0.0029 | 0.0023 | 0.1326 | 0.0370 | 0.0076 | 0.0030 | | |
| L9 | 0.0383 | 0.0015 | 0.0016 | 0.0014 | 0.0028 | 0.0056 | 0.0252 | 0.0097 | |

Tabla 1

Valores de p de la prueba multivariada de Wilks para los licores de mora

4.2 Correlación canónica

La correlación canónica es un método de análisis estadístico que nos permite encontrar la asociación entre dos grupos de variables. Para este análisis se estableció como primer bloque de variables al pH, azúcares totales, acidez y color, y como segundo bloque al contenido de polifenoles, antocianinas monoméricas, color polimérico y capacidad antioxidante. Con el empleo de este análisis multivariado, se obtuvieron dos raíces con significancia de $p < 0.05$, a continuación se muestra la raíz 1, y las variables fisicoquímicas que influyen sobre el contenido de antocianinas monoméricas y el porcentaje de color polimérico (Figura 1).

La primera raíz mostró que existe una alta correlación ($r = 0.99$) entre los parámetros fisicoquímicos pH, azúcares totales y contenido antocianinas monoméricas y color polimérico. La correlación canónica también indicó que a mayor pH hay un menor contenido de antocianinas monoméricas y mayor color polimérico, esto es decir hay un mayor oscurecimiento de las muestras.

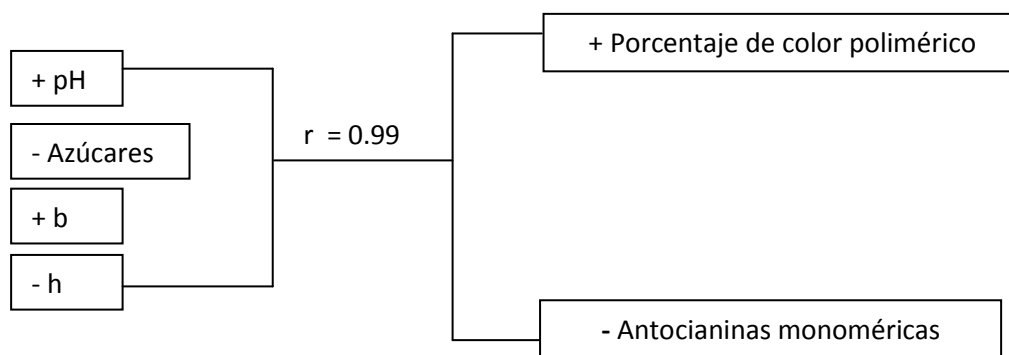


Figura 1. Raíz 1 obtenida de la correlación canónica de los licores de mora.

Sims y Morris (1985) reportaron que el pH influye es un factor importante en el oscurecimiento en los vinos durante su añejamiento. En cuanto al contenido de azúcares, en un estudio realizado por Wrolstad et al., (1990) en fresas congeladas, se encontró que el color polimérico se reducía significativamente por la adición de sacarosa, lo cual coincide con los datos obtenidos en esta investigación. Así mismo en el presente estudio, se encontró que a mayor contenido de antocianinas monoméricas hay una disminución del valor del parámetro colorimétrico b, es decir predomina el componente azul en los licores. Es sabido que a valores de pH 2 y 4, intervalo en el cual se encuentra el pH de los licores analizados, se favorece la forma quinoidal de las antocianinas, las cuales poseen un color azul (Castañeda-Ovando et al., 2009). Por otra parte, el análisis canónico mostró asociaciones entre el ángulo de matiz (h) y el contenido de antocianinas monoméricas y el porcentaje de color polimérico. Cabe mencionar que, en vinos tintos el valor del ángulo de matiz se ha relacionado con el añejamiento, siendo reportado que las antocianinas monoméricas y los pigmentos polimerizados contribuyen a la tonalidad de los vinos (Rommel et al., 1992).

En cuanto a la raíz 2, se encontró una asociación ($r = 0.97$) entre la acidez y la densidad del color, parámetro que expresa el color impartido por las antocianinas monoméricas, antocianinas copolimerizadas y de los productos de reacción del oscurecimiento, lo cual sugiere que el contenido de ácidos orgánicos influye en gran medida en el contenido de estos compuestos.

4. Conclusiones

De acuerdo a los parámetros evaluados en los licores, el análisis multivariado de la varianza indicó que existen diferencias significativas entre las muestras, lo cual puede atribuirse a las diferentes condiciones de elaboración y añejamiento. Así mismo, el análisis canónico mostró asociaciones significativas entre el pH, azúcares totales y acidez con el contenido de antocianinas monoméricas, poliméricas y la densidad del color, es decir estos parámetros fisicoquímicos son relevantes en cuanto a la presencia de compuestos bioactivos e influyen en el color de los licores.

REFERENCIAS

- AOAC. *Official Method of Analysis Association of Official Analytical Chemists*. (1997). Washington: EUA.
- Brand-Williams, W. Cuvelier, M.E., and Berset, C. (1995). Use of Free a Radical Method to Evaluate Antioxidant Activity. *Lebensmittel- Wissenschaft und-Technologie/Food Science ans Technology* 28, 25-30.
- Calvo, C. y Durán, L. (1997). Óptica y color. Temas en Tecnología de Alimentos. Instituto Politécnico Nacional. México. D. F.
- Casassa, F. and Catania, C. (2006). Pirantocianos, nuevos pigmentos en los vinos tintos: aspectos científicos e implicancias tecnológicas. *Revista Enología* 3, 40-47.
- Castañeda-Ovando, A., Pacheco-Hernández, M.L., Páez-Hernandez, M.E., Rodriguez, J.A., Galán-Vidal, C.A. (2009). Chemical studies of anthocyanins: A review. *Food Chemistry* 113, 859-871.
- Folin, C., y Ciocalteu, V. (1927). Tyrosine and Tryptophan Determination in Proteins. *Journal of Biological Chemistry* 73, 627-650.
- Giusti, M.M. and Wrolstad, R. E. (2001). Characterization and Measurement With UV-Visible Spectroscopy. In: *Current Protocols in Food Anality Chemistry*, Wrolstad, R. E., John Wiley y Sons: New York.
- Rommel A., Wrolstad R.E. and Heatherbell, D.A. (1992). balckberry juice and wine: effect of processing and storage effects on anthocyanin composition, color and appearance. *Journal of Food Science* 57, 385-391.
- Sims C. and Morris J. (1994). Effects of pH, sulfuric dioxide, storage time, and temperature on the color and stability of the red muscadine grape wine. *American Journal of Enology and Viticulture* 35, 35-39.
- Tsai, J., Huang, H. and Huang, T. (2004). Relationship between anthocyanin patterns and antioxidant capacity in mulberry wine during storage. *Journal of Food Quality* 27, 497-505.
- Wrolstad, R.E., Skrede, G., Lea, P. and Enersen, G. (1990). Influence of sugar on anthocyanin pigment stability in frozen strawberries. *Journal of Food Science* 55, 1064-1065.

Análisis de Interacción de Factores Clasificatorios de una Población en Base a las Distancias Moleculares entre sus Individuos

Bruno, Cecilia¹, Balzarini, Mónica²

¹ CONICET-Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. cebruno@agro.unc.edu.ar

² CONICET-Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. mbalzari@agro.unc.edu.ar

Introducción

Para estudiar la estructura genética de una población de individuos, es común expresar la diversidad genómica total (expresada por los distintos haplotipos moleculares) como la suma de componentes de variabilidad entre y dentro de grupos de individuos conformados por algún factor de clasificación. Generalmente, se atribuye una proporción aditiva de la variabilidad total a cada uno de los factores individuales presentes en el diseño del estudio. Así la comparación de estas componentes permite inferir la magnitud de la estructuración causada por ese factor. La estructura genética de poblaciones, en una especie, puede analizarse mediante la comparación de las frecuencias alélicas observadas dentro y entre grupos respecto a las esperadas bajo las premisas de la ley de Hardy-Weinberg la cual anula la esperanza de encontrar diferencias entre grupos (Kennington *et al.*, 2003; Hedrick, 2005). La mayoría de los métodos basados en frecuencias alélicas involucran transformaciones no lineales de los datos genéticos y son válidos sólo bajo una serie de supuestos que deben realizarse respecto a los procesos evolutivos subyacentes. Por el contrario, la información sobre divergencia a nivel molecular procesada en el formato de una partición de suma de cuadrados si bien depende de un modelo, como el del ANAVA, demanda menos supuestos biológicos.

Dada la naturaleza booleana multidimensional de la información provista por los haplotipos moleculares, el análisis es esencialmente multivariado. Debido a la dimensionalidad de los datos genómicos, las sumas de cuadrados (SC) para el análisis de las fuentes de variación se obtienen a partir de las distancias entre los pares de muestras individuales y no a partir de los datos originales provistos por los marcadores moleculares. Debido a relaciones entre SC y sumas de distancias al cuadrado, la SC asociada con cualquier término de un modelo lineal puede ser calculada directamente a partir de una matriz de distancias (Gower, 1966; Li, 1979). Estos cálculos se construyen generalmente desde la matriz de distancias Euclídeas (al cuadrado) (aunque en algunas aplicaciones se han utilizado medidas de distancias alternativas) y se usan para contrastar hipótesis sobre variabilidad entre y dentro de grupos en estructuras jerárquicas (anidadas) de factores. Sin embargo, existen situaciones donde los factores producen clasificaciones cruzadas de los individuos y por tanto es de interés probar la significancia de la interacción entre ellos. El objetivo de este trabajo es proponer y evaluar una prueba estadística para la interacción entre factores en el caso de diseños a dos vías de clasificación con estructura ortogonal, FIDA (*Factor Interaction from Distance Analysis*) e ilustrar la aplicación de FIDA sobre distintos conjuntos de datos moleculares simulados para poner de manifiesto la potencia de la prueba.

Distancias en el análisis de varianza molecular

Supongamos que para cada grupo existen n muestras de haplotipos tratadas en laboratorio con m marcadores. Consecuentemente, se tendrán n observaciones con un número m de variables. Si los marcadores se expresan como variables binarias, para cada muestra es posible conformar una observación multivariada (m -dimensional) que lleva valores 1 o 0 para cada uno de los marcadores según el marcador esté presente o ausente en la muestra. El vector booleano m -dimensional es denotado como $\mathbf{p}' = [p_1, p_2, \dots, p_m]$ donde $p_i = 1$ con $i=1, \dots, m$ si el marcador m está presente y cero si está ausente. La diferencia entre dos muestras y_j y y_k es definida como $\mathbf{p}_j - \mathbf{p}_k$. Se define una métrica de distancia Euclídea

entre las muestras y_j y y_k como $d_{jk}^2 = (\mathbf{p}_j - \mathbf{p}_k)' \mathbf{W} (\mathbf{p}_j - \mathbf{p}_k)$ donde \mathbf{W} es una matriz de pesos que pueden ser diferenciables para los distintos marcadores. Si todos los marcadores se asumen independientes e igualmente informativos entonces $\mathbf{W} = \mathbf{I}$ y la métrica de distancia es igual al número de diferencias entre las dos muestras, *i.e.* complemento a uno del índice de similitud Emparejamiento Simple (Apostol, 1993).

Una distancia comúnmente usada es la distancia de Excoffier (Excoffier *et al.*, 1992), en ésta las disimilitudes o faltas de coincidencia son expresadas como fracción del número total de marcadores que participan en la comparación de cada par de perfiles. Si todos los individuos que se comparan son genotipados con el mismo número de marcadores entonces esta distancia es sólo un múltiplo del complemento a 1 del índice de Emparejamiento Simple. Las metodologías de análisis basadas en sumas de cuadrados de las distancias no depende de la forma particular de \mathbf{W} que se elija siempre que las distancias obtenidas constituyan una métrica y por tanto cumplan con las condiciones que definen a éstas.

Prueba de interacción de factores desde matrices de distancias

A continuación se presenta una prueba estadística desarrollada para evaluar la significancia de la interacción $A \times B$ en cualquier diseño con dos factores cruzados para cualquier número de niveles de cada factor, que en adelante llamaremos FIDA (*Factor Interaction from Distance Analysis*). La prueba no demanda la elección de un método de permutación para hallar la significancia y goza de las propiedades del *test* no paramétrico de Kruskal-Wallis (Conover, 1999) ya que se basa en la aplicación del mismo sobre el valor absoluto de residuos de distancias entre y dentro los grupos formados por ambos factores. El algoritmo comprende: cálculo de la matriz de distancia y partición de la misma en bloques de distinto tipos de distancia, posterior cálculo de residuos y análisis de varianza no-paramétrico sobre esos residuos.

La matriz de distancias entre pares de individuos se particiona de manera tal que se identifiquen los siguientes bloques de distancias: 1) Bloque de distancias “Dentro”: comprende inter-distancias entre individuos de un mismo grupo, por tanto existen tantos bloques de este tipo como grupos (combinaciones de factores) haya; 2) Bloque de distancias “Entre factor B dentro de factor A ”, existen tantos bloques como niveles del factor A ; 3) Bloque de distancias “Entre factor A dentro de factor B ” existen tantos bloques como niveles del factor B y 4) Bloque de distancias “Entre factor A y entre factor B ”, con tantos bloques como $(B-1)B$ (Fig. 1).

| Factor A | | A_1 | | A_2 | |
|----------|----------|-------|-------|-------|-------|
| | Factor B | B_1 | B_2 | B_1 | B_2 |
| A_1 | B_1 | B1 | | | |
| | B_2 | B2 | B1 | | |
| A_2 | B_1 | B3 | B4 | B1 | |
| | B_2 | B4 | B3 | B2 | B1 |

Fig. 1. Esquema de los bloques conformados por la partición de la matriz de distancia.

Nota: B1: Bloques de distancia “Dentro”, B2: Bloque de distancias entre factor $B[A]$, B3; Bloque de distancias entre $A[B]$, B4: Bloque de distancia entre factor A y entre factor B .

Se obtienen las distancias promedios para cada bloque de distancias (B1 a B4), luego se obtiene el valor absoluto de la diferencia entre una distancia del bloque y la media de las distancias del mismo bloque. Sobre el valor absoluto de la diferencia de cada distancia respecto a su media, se ajusta un análisis de varianza no-paramétrico de Kruskal Wallis. A través del análisis de varianza no-paramétrico se evalúa la hipótesis de homogeneidad de varianzas dentro de los distintos bloques. El valor promedio de la variable de análisis (rangos de los valores absolutos de los residuos) será el mismo para todos los bloques sólo cuando la variabilidad dentro de cada uno de los cuatro bloques sea la misma. Si los valores esperados en los bloques que contiene distancias “entre” (B2, B3, B4) no difieren significativamente de los valores esperados en el bloque B1 (que contiene distancias “dentro” de un grupo de individuos o entre individuos de un mismo grupo) entonces no existen evidencia para rechazar la hipótesis nula y suponer que haya interacción entre los factores. La interacción se produce cuando las diferencias entre los niveles de un factor dentro de un mismo nivel de un segundo factor, son distintas a las diferencias obtenidas para otros niveles del segundo factor. Si esto ocurre (*i.e* si hay interacción entre factores) los valores absolutos de los residuos provenientes de distancias que involucran perfiles de distintos niveles de uno o ambos factores (Bloques B2, B3 y B4), serán mayores que aquellos provenientes de distancias del bloque B1, donde sólo se estima la variabilidad residual y no la debida a la interacción. Para evitar supuestos distribucionales en la variable de análisis (valor absoluto de los residuos entre distancias observadas y distancias esperada sin interacción) se ajusta una prueba no paramétrica, basada en rango, por lo que el valor de significancia es obtenido como en Kruskal Wallis (Conover, 1999)

Materiales y Métodos

En el FIDA la hipótesis nula (H_0) que se somete a prueba es que no hay interacción entre los factores. Como contraste, la hipótesis alternativa H_1 es que la interacción es diferente de cero. Con una prueba estadística no-paramétrica, obtenida por métodos de permutación, se acepta o no la hipótesis nula. Al rechazar H_0 cuando es verdadera, se estaría incurriendo en un error, el cual se denomina Error de tipo I, cuya probabilidad es usualmente denotada por α o *nivel de significación* de la prueba. También se puede aceptar H_0 cuando es falsa, en este caso se comete el Error de tipo II cuya probabilidad de ocurrencia se denomina β . La cantidad $(1 - \beta)$ se llama *potencia de la prueba* y es la probabilidad de la prueba de detectar verdaderas diferencias, por lo cuál depende de las magnitudes de las verdaderas diferencias que se estén produciendo entre los bloques de análisis. Para conocer el Error de tipo I y la magnitud del Error de tipo II de la propuesta de análisis de varianza molecular no

jerárquico propuesta, se evaluaron empíricamente, a través de simulaciones, los errores de Tipo I y Tipo II.

En la evaluación de las propiedades de FIDA, se partió de dos situaciones hipotéticas para una matriz $n \times m$ donde $n=24$ y $m=20$. El diseño de experimentos fue un bifactorial con dos niveles cada factor. Las situaciones hipotéticas fueron construidas bajo: (I) hipótesis nula falsa, es decir que al menos una de las diferencias entre los bloques definidos sobre la matriz de distancias es distinta de cero, *i.e.*, la interacción es estadísticamente significativa y II) hipótesis nula verdadera, es decir que no hay interacción entre los factores.

En cada una de las situaciones (I) y (II), se simularon 100 bases de datos a partir de una perturbación introducida a través de una distribución binomial con parámetros $n=1$ y $p=0.20$. Sobre cada una de las cien nuevas matrices de datos binarios creados para cada una de las situaciones, se calcularon tres medidas de distancia entre todos los pares de individuos: la distancia de Excoffier, el complemento a uno del índice de emparejamiento simple y la distancia de Bray-Curtis. Se calcularon las distancias promedios para cada bloque de distancias a partir del cual se reagruparon los datos, según el esquema presentado en Fig. 3.3. Sobre el valor absoluto de la diferencia de la distancia entre todos los pares de individuos y la media del bloque al que pertenece según el re-agrupamiento, se ajustó un análisis de la varianza no-paramétrico de Kruskal Wallis.

A través de las cien corridas o ajustes del análisis propuesto y bajo hipótesis nula verdadera, se contó la cantidad de veces que la prueba no aceptó la hipótesis es decir donde se rechaza una hipótesis verdadera que suponía la no existencia de interacción. Este valor estimado empíricamente representa el “tamaño de la prueba”.

Para estimar, al menos empíricamente, la potencia de la prueba se simularon para la situación (I) diferentes niveles de interacción: (I.a) alta interacción, donde las diferencias entre los individuos de los niveles del factor B para un mismo nivel del factor A eran altas respecto a las diferencias entre los niveles de los factores B con el otro nivel de A donde esas diferencias eran prácticamente nulas. (I.b) media interacción, donde la diferencia entre individuos de los niveles de B que se encuentran bajo un mismo nivel del factor A presentan una diferencia máxima en el 60% de sus loci y (I.c) baja interacción; la diferencia máxima se da solo en un 30% de los loci. Para conocer la potencia de la prueba, se contó en cada situación la cantidad de veces que el valor p del análisis de la varianza no-paramétrico arrojó valores por debajo del nivel de significación ($\alpha=0.05$).

El método FIDA fue comparado con los resultados arrojados por un software que estima la interacción entre los niveles de dos o más factores pero sólo en situaciones donde los datos son balanceados, *i.e.*, todos los niveles de los factores tienen la misma cantidad de datos y no falta ningún nivel.

Resultados y discusión

En la Tabla 1 se presenta para cada situación simulada (Ia., Ib, Ic y II) las tasas de error estimadas para la prueba propuesta en esta tesis y para la prueba de interacción implementada en el software PERMANOVA v.1.6 (Anderson y ter Braak, 2003). Ambas

fueron aplicadas teniendo como *input* una matriz de distancias entre individuos de dimensión 24×24 y calculada en base a la métrica de Excoffier.

Tabla 1. Tasas de Error Tipo I y Tipo II obtenidas por simulación para la prueba de hipótesis de la interacción, bajo situaciones de alta, mediana, baja y nula interacción, de los procedimientos FIDA y PERMANOVA con el método de permutación de filas.

| Situación | Interacción | FIDA | | PERMANOVA | |
|-----------|--------------|----------------------|-----------------------|----------------------|-----------------------|
| | | Tasa de Error Tipo I | Tasa de Error Tipo II | Tasa de Error Tipo I | Tasa de Error Tipo II |
| I | Alta | Nc | 0.05 | Nc | 0 |
| | Media | Nc | 0.20 | Nc | 0 |
| | Baja | Nc | 0.28 | Nc | 0.15 |
| II | Nula | 0.097 | Nc | 0.05 | Nc |

Nota: Nc: no corresponde, por definición el Error de Tipo I se calcula bajo hipótesis nula cierta y el Error de Tipo II bajo hipótesis nula falsa.

Los resultados en Tabla 1 sugieren que la prueba FIDA presenta una alta potencia para detectar el efecto de interacción entre los factores, aún cuando este efecto es bajo. Esta potencia es muy similar a la que presenta PERMANOVA v.1.6 a través de 4999 corridas de permutación de filas. La ventaja de FIDA es que no necesita que el diseño de experimento tenga el mismo número de repeticiones para cada nivel de factores, mientras que la versión 1.6 de PERMANOVA sólo puede ser implementada para diseños balanceados, *i.e.*, deben haber el mismo número de datos para cada combinación de factores y no puede faltar ninguna combinación. A diferencia de PERMANOVA, FIDA, es una prueba específica de interacción y no del efecto de factores principales. Sin embargo, en muchas situaciones prácticas el énfasis está puesto en conocer si la interacción entre los factores es estadísticamente significativa. Además, si la interacción es estadísticamente significativa, no se recomienda el *test* de los efectos principales. En los diseños con factores cruzados se recomienda analizar primero las interacciones y luego los efectos principales siempre y cuando la interacción no haya sido estadísticamente significativa ya que una interacción alta podría enmascarar efectos principales de uno o ambos factores.

REFERENCIAS

- Anderson M, ter Braak T (2003) Permutation test for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation*. 73(2):85-113
- Apostol B-L, Black W-C, Miller B-R, Reiter P, Beaty B-J (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. *Theoretical and Applied Genetics* 86: 991-1000.
- Conover W-J (1999) *Practical nonparametric statistics*, 3rd edition. New York: John Wiley & Sons.
- Excoffier L, Smouse P-E, Quattro J-M (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Gower J-C (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53(3 and 4):325-338.
- Hedrick P (2005) *Genetic of Populations*. Arizona State University. Jones and Bartlett publishers. Third Edition.

Kennington W-J, Gockel J, Partridge L (2003) Testing for Asymmetrical Gene Flow in a *Drosophila melanogaster* Body-Size Cline. *Genetics* 165: 667-673.
Li C-C (1976) *Population Genetics*. Boxwood, Pacific Grove, Calif.

Relación del Ángulo Interdental con el Plano Mento-Nasal (Plano Estético de Ricketts)

Gustavo Alfonso Sotelo y Soto¹, Noralia Ramírez Chávez², Erika Sotelo Acosta³,
Juan José Trujillo Fandiño⁴, Julia Aguirre Sampieri⁵, Jorge de Jesús Serena Melo⁶,
Minerva Reyes Felix⁷

¹Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. gsotelo@uv.mx

²Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. nramirez@uv.mx

³Fac. de Medicina. Universidad Veracruzana. Xalapa, Ver. zs07017321@malum1.uv.mx

⁴Hospital Juárez. S. S .A. México, D. F. jjtrujillo@yahoo.mx

⁵Fac. de Odontología. UNAM. México, D. F. lurue19@hotmail.com

⁶Fac. de Odontología. Universidad Veracruzana. Xalapa, Ver. jserena@uv.mx

⁷Fac. de Estadística. Universidad Veracruzana. Xalapa, Ver. minreyes@uv.mx

Esta investigación se ubica en un contexto universal en cuanto a las medidas faciales que se indagan, ya que es una relación existente en todos los seres vivos dentados y especialmente en los humanos, pero en este caso corresponde a una muestra tomada a pacientes que acuden a una consulta particular de especialidad dental. La muestra es representativa del total de expedientes clínicos de pacientes atendidos en este consultorio y corresponde al 20 por ciento, del cual la mitad son del sexo femenino y la restante del sexo masculino.

Las mediciones faciales a que se hace referencia son el ángulo formado por los dientes incisivos centrales superiores e inferiores y la distancia del labio inferior al plano mento-nasal, (plano estético de Ricketts). se obtuvieron de las radiografías extraorales laterales de cráneo para estudio cefalométrico, capturadas de cada uno de los pacientes que aquí representan el objeto de estudio. Dicha medición tiene su importancia en la oclusión dental funcional en armonía con la cara y fue estudiada y dada a conocer a la Odontología por Robert Murray Ricketts alrededor de los años 50 y desde entonces sus seguidores han profundizado en esa línea de investigación, que permite conocer la necesidad de modificar la oclusión o función dental, con cambios ortopédicos, ortodónticos, quirúrgicos o una combinación de ellos. Al mismo tiempo que se conoce y estudia la oclusión dental funcional para saber si será necesario modificarla, con la finalidad de favorecer una buena función masticatoria, fonética, muscular y por ende digestiva, se establece la necesidad de realizar cambios faciales en función de una armonía y estética facial, aspecto que desde siempre ha sido importante para las relaciones humanas, ya que la belleza se considera hoy en día un requisito más para el éxito.

Metodología. La investigación que aquí se presenta, consta de tres variables dos de tipo escala de razón que corresponden a la edad y el ángulo la relación del ángulo interdental con el plano mento-nasal; una escala de intervalo que es la medida Ricketts Se busca conocer estadísticamente el porcentaje de pacientes cuyas medidas cumplen esta premisa. Finalmente se presenta una escala nominal que corresponde a los géneros femenino y masculino. La variable edad de los pacientes objeto de estudio se ubica en el rango de los nueve a los veinticinco años, con una temporalidad actual y una ubicación en la región central del Estado de Veracruz.

| VARIABLES | Total de casos | Media | Mediana | Moda | Frecuencia | Valor mínimo | Valor máximo | Desv. Estándar |
|---------------------|----------------|----------|----------|----------|------------|--------------|--------------|----------------|
| Edad | 100 | 15.5600 | 15.0000 | 12.00000 | 13 | 9.00000 | 25.0000 | 4.22910 |
| Ángulo interincisal | 100 | 122.4250 | 121.2500 | 128.0000 | 6 | 99.00000 | 158.5000 | 11.41799 |
| Ricketts | 100 | 0.5300 | 0.2500 | 0.000000 | 15 | -9.50000 | 7.5000 | 2.65130 |

De acuerdo a la tabla de estadísticas descriptivas, se puede observar que la edad promedio de las personas del estudio es de 15.56 años de edad y la variabilidad de los datos es de 4.23 años, también se presenta que el valor mínimo registrado de edad es de 9 años y la edad máxima encontrada es de 25 años.

Respecto a la variable del ángulo interincisal se tiene que el valor promedio del ángulo de los pacientes es de 121.25 grados con una desviación estándar de 11.42 grados, además se encuentra que el valor mínimo del ángulo de los pacientes es de 99 grados y el valor máximo es de 158.5 grados.

| | Frecuencia | Frecuencia Acumulada | % | % acumulado |
|------|------------|----------------------|----|-------------|
| -9.5 | 1 | 1 | 1 | 1 |
| -5.5 | 1 | 2 | 1 | 2 |
| -4.5 | 1 | 3 | 1 | 3 |
| -4 | 1 | 4 | 1 | 4 |
| -3.5 | 1 | 5 | 1 | 5 |
| -3 | 1 | 6 | 1 | 6 |
| -2 | 13 | 19 | 13 | 19 |
| -1.5 | 7 | 26 | 7 | 26 |
| -1 | 5 | 31 | 5 | 31 |
| -.50 | 4 | 35 | 4 | 35 |
| 0 | 15 | 50 | 15 | 50 |
| .50 | 4 | 54 | 4 | 54 |
| 1 | 10 | 64 | 10 | 64 |
| 1.5 | 5 | 69 | 5 | 69 |
| 2 | 9 | 78 | 9 | 78 |
| 2.5 | 4 | 82 | 4 | 82 |
| 3 | 7 | 89 | 7 | 89 |
| 3.5 | 2 | 91 | 2 | 91 |
| 4 | 1 | 92 | 1 | 92 |
| 4.5 | 1 | 93 | 1 | 93 |
| 5 | 3 | 96 | 3 | 96 |
| 6 | 1 | 97 | 1 | 97 |
| 7 | 2 | 99 | 2 | 99 |
| 7.5 | 1 | 100 | 1 | 100 |

Por otra parte los resultados de la variable Ricketts que se refiere a la oclusión dental reporta una valor promedio de 0.53mm, un valor central de los datos de 0.25mm y una moda con una frecuencia de 15 pacientes que presentan el valor de cero. Por otra parte se encuentra que el menor valor de oclusión encontrado es de -9.5mm y el mayor es de 7.5mm. En seguida se presenta una tabla de frecuencia de la variable Ricketts, que muestra cada uno de los valores de oclusión en los pacientes, así como su respectiva frecuencia y el

porcentaje de pacientes que registran dichos valores. Se puede destacar que los valores de oclusión que presentan mayor frecuencia son el valor de 0 con una frecuencia de 15 pacientes, el valor de -2 con una frecuencia de 13 y el valor de 1 con una frecuencia de 10 pacientes.

ESTADÍSTICAS DESCRIPTIVAS POR GÉNERO

| | Frecuencia | Frecuencia Acumulada | Porcentaje | Porcentaje acumulado |
|------|------------|----------------------|------------|----------------------|
| -9.5 | 1 | 1 | 2 | 2 |
| -4 | 1 | 2 | 2 | 4 |
| -3.5 | 1 | 3 | 2 | 6 |
| -2 | 5 | 8 | 10 | 16 |
| -1.5 | 2 | 10 | 4 | 20 |
| -1 | 3 | 13 | 6 | 26 |
| -.50 | 1 | 14 | 2 | 28 |
| 0 | 8 | 22 | 16 | 44 |
| .50 | 1 | 23 | 2 | 46 |
| 1 | 7 | 30 | 14 | 60 |
| 1.5 | 3 | 33 | 6 | 66 |
| 2 | 3 | 36 | 6 | 72 |
| 2.5 | 4 | 40 | 8 | 80 |
| 3 | 4 | 44 | 8 | 88 |
| 3.5 | 1 | 45 | 2 | 90 |
| 4 | 1 | 46 | 2 | 92 |
| 4.5 | 1 | 47 | 2 | 94 |
| 5 | 1 | 48 | 2 | 96 |
| 7 | 1 | 49 | 2 | 98 |
| 7.5 | 1 | 50 | 2 | 100 |

| Variables | Total de casos | Media | Mediana | Moda | Frecuencia | Valor mínimo | Valor máximo | Desv. Estándar |
|---------------------|----------------|----------|----------|--------------|------------|--------------|--------------|----------------|
| Edad | 50 | 15.1200 | 14.5000 | 13.0000 0 | 6 | 9.0000 | 25.0000 | 4.023604 |
| Angulo interincisal | 50 | 119.9300 | 120.0000 | Múltiple | 3 | 100.0000 | 148.0000 | 9.338318 |
| Ricketts | 50 | 0.7600 | 1.0000 | 0.00000 0 | 8 | -9.5000 | 7.5000 | 2.796207 |

La edad promedio para el género masculino es de 15.12 años con una desviación estándar de 4.02 años. La edad mínima es de 9 años y la edad máxima es de 25 años.

El ángulo interincisal en promedio es de 119.93 grados con una desviación estándar de 9.33 grados, además presenta una moda múltiple y los valores mínimo y máximo del ángulo son de 100 y 148 grados respectivamente. En lo referente a la medida de oclusión dental se observa que las pacientes tienen en promedio una medida de 0.76mm con una desviación estándar de 2.79mm, así mismo se tiene que el valor máximo de oclusión dental es de 7.5mm y el valor mínimo de -9.5mm. También se reporta que la moda se da en el valor de 0mm con una frecuencia de 8 pacientes. En lo que respecta a la tabla de frecuencias se

observa que además de la moda antes mencionada, también se tiene que la medida de oclusión de 1mm presenta una frecuencia de 7.

| Estadísticas descriptivas del género femenino | | | | | | | | |
|---|----------------|----------|----------|----------|------------|--------------|--------------|----------------|
| VARIABLES | Total de casos | Media | Mediana | Moda | Frecuencia | Valor mínimo | Valor máximo | Desv. Estándar |
| Edad | 50 | 16.0000 | 16.0000 | 12.00000 | 8 | 9.00000 | 25.0000 | 4.42165 |
| Ángulo interincisal | 50 | 124.9200 | 123.2500 | Múltiple | 3 | 99.00000 | 158.5000 | 12.78646 |
| Ricketts | 50 | 0.3000 | 0.0000 | -2.00000 | 8 | -5.50000 | 7.0000 | 2.50510 |

| Tabla de frecuencia para el género femenino de la variable ricketts | | | | |
|---|------------|----------------------|------------|----------------------|
| | Frecuencia | Frecuencia Acumulada | Porcentaje | Porcentaje acumulado |
| -5.5 | 1 | 1 | 2 | 2 |
| -4.5 | 1 | 2 | 2 | 4 |
| -3 | 1 | 3 | 2 | 6 |
| -2 | 8 | 11 | 16 | 22 |
| -1.5 | 5 | 16 | 10 | 32 |
| -1 | 2 | 18 | 4 | 36 |
| -.50 | 3 | 21 | 6 | 42 |
| 0 | 7 | 28 | 14 | 56 |
| .50 | 3 | 31 | 6 | 62 |
| 1 | 3 | 34 | 6 | 68 |
| 1.5 | 2 | 36 | 4 | 72 |
| 2 | 6 | 42 | 12 | 84 |
| 3 | 3 | 45 | 6 | 90 |
| 3.5 | 1 | 46 | 2 | 92 |
| 5 | 2 | 48 | 4 | 96 |
| 6 | 1 | 49 | 2 | 98 |
| 7 | 1 | 50 | 2 | 100 |

En resumen se tiene que la edad promedio para el caso del género femenino es de 16 años con una desviación estándar de 4.42 años, además se mantiene que la edad mínima es de 9 años y la edad máxima es de 25 años. Con respecto al ángulo interincisal, en promedio se obtuvo un valor de 124.92 grados con una desviación estándar de 12.7 grados. En lo referente a la medida de oclusión dental se observa que los pacientes del sexo femenino tienen en promedio medidas de 0.3mm y una desviación estándar de 2.50mm, así mismo un valor máximo de 7 un valor mínimo de -5.5mm de oclusión. También se reporta que la moda se da en el valor de -2mm con una frecuencia de 8 pacientes.

De la tabla de frecuencias podemos observar los valores de oclusión para el género femenino, siendo los más considerables los valores indicados en la moda y la medida de oclusión de 0 y 2mm con una frecuencia de 7 y 6 pacientes respectivamente.

En el diagrama de cajas y alambres se observa que el género masculino presenta valores de oclusión mayores que el del género femenino, de la misma forma la mediana del género masculino (1mm) es mayor que la del femenino (0mm). Por otra parte la variabilidad de los datos que presenta el género femenino es mayor que el masculino. También cabe mencionar que hay un valor outlier con valor de -9.5mm en el género masculino.

El diagrama de cajas y alambres de la variable ángulo interincisal muestra que para el género femenino se presentan los valores más altos en las medidas de ángulo a diferencia del género masculino, aunque cabe mencionar que hay mayor variabilidad en las mediciones registradas para las mujeres. Otro aspecto que se destaca es que en ambos casos se presentan datos outliers, es decir que se alejan de los demás mediciones y que toman valores para el caso del género femenino mayores a 150 grados, caso contrario al género masculino que registra su valor outlier menor a 150.

Además el 75% de las pacientes presenta valores por debajo de 130 grados de ángulo interincisal, en cambio en los pacientes del género masculino, el 75% de pacientes registra valores por debajo de 125 grados.

Estadísticas descriptivas del rango de edad de 9 a 17 años

| | Valid N | Mean | Median | Mode | Frequency | Minimum | Maximum | Std. Dev. |
|---------------------|---------|----------|----------|----------|-----------|----------|----------|-----------|
| edad | 69 | 13.2319 | 13.0000 | 12.00000 | 13 | 9.00000 | 17.0000 | 2.44435 |
| angulo interincisal | 69 | 123.5507 | 123.0000 | 128.0000 | 5 | 99.00000 | 158.5000 | 12.17946 |
| Ricketts | 69 | 0.6594 | 0.5000 | 0.000000 | 11 | -5.50000 | 7.5000 | 2.36471 |

Estadísticas descriptivas del rango de edad de 18 a 25 años

| | Valid N | Mean | Median | Mode | Frequency | Minimum | Maximum | Std.Dev. |
|---------------------|---------|----------|----------|----------|-----------|----------|----------|----------|
| edad | 31 | 20.7419 | 20.0000 | 18.00000 | 7 | 18.0000 | 25.0000 | 2.294453 |
| angulo interincisal | 31 | 119.9194 | 118.5000 | 115.0000 | 4 | 103.5000 | 141.0000 | 9.197007 |
| Ricketts | 31 | 0.2419 | 0.0000 | 0.000000 | 4 | -9.5000 | 7.0000 | 3.224570 |

Referencias

- Bellini, F.(2004) Ortodoncia diagnóstico y planificación clínica. 2ª. Edición Editora Artes Médicas Ltda.
- Canut, J.(2000) Ortodoncia Clínica y Terapéutica. 2ª. Edición. Editorial Masson.
- Ferring, V., Pancherz, H. “*Divine proportion in the growing face*” Am J Orthod Dentofacial Orthop 2008;134:472-9.
- Kiekens,R.M.A., Kuijpers-J.A. M.,et. al. En “*Putative golden proportions as predictorof facial esthetics in adolescents* “.Am J Orthod Dentofacial Orthop 2008;134:480-3.

Alometría Interespecífica de Árboles de la Selva Alta Perennifolia en el Volcán San Martín Pajapan, Veracruz

Báez-Hernández. A¹., A. J. Martínez²., S.M. Vázquez–Torres³ & J. G. Montero–Mora⁴

¹ Instituto de Investigaciones Biológicas. Universidad Veracruzana. arturo_bh27@hotmail.com

² Instituto de Neuroetología. Universidad Veracruzana. amartinez@uv.mx

³ Instituto de Investigaciones Biológicas. Universidad Veracruzana. savazquez@uv.mx

⁴ Facultad de Estadística. Universidad Veracruzana. jmontero@uv.mx

Resumen.

La forma de crecimiento de los árboles tiene un amplio significado ecológico y la ganancia en altura afecta la interceptación de luz, estabilidad biomecánica y determina parte de la historia de vida de cada especie. De ahí que la alometría sea un buen método comparativo para analizar los atributos arquitectónicos de los árboles. En el presente estudio se obtuvo la alometría para tres atributos de tamaño de diecinueve especies de árboles de la selva alta perennifolia en el volcán San Martín Pajapan, Veracruz. Se midió la altura, diámetro a la altura del pecho y cobertura de los árboles en dos franjas de cuatro hectáreas dentro del bosque lluvioso tropical. Con los atributos de tamaño se obtuvieron las distribuciones de frecuencias y se realizaron comparaciones univariadas y análisis alométricos a partir de modelos de regresión simple. La altura y cobertura de las diecinueve especies de árboles muestran diferentes trayectorias alométricas y esto es reflejo de la variabilidad interespecífica y de la plasticidad al gradiente lumínico dentro del dosel de este tipo de bosque tropical, así como a las estrategias de crecimiento de cada especie. Asimismo, la variación en la estatura (A_{max}) se asocia positivamente ($r^2 = 0.67$) con la cobertura de la copa (C_{max}). Sin embargo, esto sólo explica parte de las diferencias funcionales entre las especies en cuanto a los costos y beneficios de inversión a crecimiento vertical y horizontal debido a la gran variedad de alturas y coberturas dentro de las especies. Por lo tanto, la arquitectura del árbol que se refiere a la forma y la posición espacial de los atributos de crecimiento determine en parte el medio más eficiente para la distribución del follaje y acceso al dosel superior, así como soporte biomecánico para tolerar la presión por estrés debido al viento u otra tensión ambiental.

Palabras clave: alometría, árboles, asociación, interespecífica, selva.

Introducción.

Los bosques templados y selvas tropicales se caracterizan por su buen desarrollo arquitectónico del dosel, debido a las múltiples capas de copas y al crecimiento vertical de cada especie, lo cual permite que cohabiten (Richards 1996; Kohyama 1992, 1993; Takahashi *et al.*, 2001). Es así que la forma de crecimiento de cada especie expresa un amplio significado ecológico (Küppers 1989, Archibal y Bond 2003), ya que el incremento en altura es consecuencia de la presión de diversos factores abióticos (Niklas 1994), de ahí que el diseño arquitectónico del árbol sea producto de escenarios de costo-beneficio impuestos por los diferentes requerimientos para el desarrollo de las plantas. El tamaño de los árboles está estrechamente relacionado con la dinámica poblacional, ya que una alta densidad de follaje en el dosel superior puede suprimir el crecimiento y el reclutamiento de

nuevos individuos (Aiba y Kohyama 1996). Es evidente que las propiedades ecológicas de las comunidades vegetales se pueden representar a partir de interrelacionar atributos dependientes del tamaño usando la alometría, puesto que esta aproximación define los cambios relativos a las dimensiones de las estructuras dentro de los organismos que están correlacionados con el tamaño global (Niklas 1994; Gayon 2000). Se ha demostrado que el tamaño máximo de cada especie está definido o relacionado con la forma o arquitectura del árbol a nivel intraespecífico (Archibal y Bond 2003, Martínez y López Portillo 2003, Méndez-Alonzo y López-Portillo 2008) e interespecífico (King 1990, Sterck y Bongers 1998, Kohyama *et al.*, 2003, Poorter *et al.*, 2005, Poorter *et al.*, 2006, Poorter *et al.*, 2008). Es así que se han realizado diversos estudios alométricos en distintas especies de árboles que habitan los bosques húmedos tropicales y templados (Kohyama 1987; Kohyama y Hotta 1990; Takahashi *et al* 2001; Turner *et al.*, 1995; Takahashi y Kohyama 1997; López-Serrano *et al.*, 2005, Akashi 2006). Así se describen las distribuciones de frecuencias de los tres atributos de tamaño, se estiman los parámetros dasométricos y las relaciones alométricas interespecíficas entre el diámetro a la altura del pecho *vs.* la altura y la cobertura.

Metodología

Sitio de estudio.

Se realizó en el volcán San Marín Pajapan donde esta presente la selva alta perennifolia entre el año 1995 y el 2000, con 1250 msnm dentro del complejo montañoso de los Tuxtlas, Veracruz, y de la planicie costera del golfo de México, el sitio se encuentra inmerso dentro de la vertiente noroeste del volcán San Martín Pajapan a una altitud de 600 m.

Diseño del muestreo.

Se realizó la selección del sitio en base a las condiciones fisonómicas, una vez seleccionados los sitios, se ubicaron y delimitaron dos franjas de 400 x 100 m y así se obtuvieron dos áreas de cuatro hectáreas cada una. La distancia entre cada sitio fue de 2 Km. entre franja y franja. Posteriormente se delimitaron cuadros de 100 x 100 m y después se realizó el censo y registro de los árboles perennes que se distribuían en cada área. Con lo cual se procedió al registro de los siguientes atributos de tamaño: diámetro a la altura del pecho (DAP) (≥ 20 cm), altura total del árbol (ALT) (m) y cobertura de la copa del árbol (COB) (m^2). Esta última variable se calculo a partir del a formula de la elipse ($0.25 \cdot \pi \cdot DM \cdot dm$), donde *DM*, representa el diámetro mayor y *dm* el diámetro menor. Para la realización de este estudio, se seleccionaron diecinueve especies de árboles de dieciocho familias diferentes debido a que fueron las más abundantes en ambos sitios: *Alchornea latifolia*, *Calatola laevigata*, *Clethra microphylla*, *Dendropanax arboreus*, *Guarea glabra*, *Heliocarpus appendiculatus*, *Ilex discolor*, *Inga quaternata*, *Ocotea dendrodaphne*, *Pouteria lucentifolia*, *Pouteria torta*, *Pseudolmedia oxyphyllaria*, *Psychotria chiapensis*, *Rhedia edulis*, *Roupala montana*, *Sterculia mexicana*, *Turpinia occidentalis*, *Virola guatemalensis* y *Zuelania guidonia*. Se utilizó estadística descriptiva y análisis exploratorios para obtener las categorías de tamaño de los atributos de crecimiento, se obtuvo el registró de los valores promedio y errores estándar. Se calculo la altura máxima (A_{max}) y cobertura máxima (C_{max}), a partir del percentil 95 para cada especie de árbol. Los atributos de crecimiento de los árboles fueron comparados entre especies utilizando pruebas de ANDEVA de una vía para determinar diferencias significativas entre especies y las tres variables de crecimiento. Asimismo, para determinar los coeficientes alométricos de las

diecinueve especies de árboles se obtuvieron las relaciones alométricas entre: diámetro a la altura del pecho (cm) vs. altura total del árbol (m), y diámetro a la altura del pecho vs. cobertura de la copa (m²). Esto se realizó aplicando el modelo de regresión lineal: $y = bx^k$ donde y es la magnitud de la diferencia en crecimiento para las variables de repuesta, x representa el tamaño del soma, k la constante de diferenciación crecimiento-proporción y b constante (ordenada al origen). Los análisis fueron procesados en los paquetes estadísticos SigmaPlot 10.0 Copyright © 2006 Systat Software, Inc. Y Statistica 7 Copyright © 2004 Statsoft. Para no violar los supuestos de normalidad y homogeneidad de varianzas los datos fueron transformados a logaritmo base 10.

Resultados.

Los parámetros dasométricos estimados de las diecinueve especies de árboles presentan amplia variabilidad. Solo tres especies *Pseudolmedia oxyphyllaria* (12.2), *Rheedia edulis* (11.4) y *Sterculia mexicana* (9.4) son las que presentan la más alta densidad encontrando para este parámetro dos grupos más. En tanto que la altura máxima estimada para los árboles indica cuatro grupos en cuanto al tamaño, ya que las especies *Pseudolmedia oxyphyllaria*, *Psychotria chiapensis*, *Calatola laevigata*, *Guarea glabra* y *Turpinia occidentalis* son las de menor A_{max} y C_{max} , y en cambio las especies *Ilex discolor*, *Virola guatemalensis*, *Clethra macrophylla* y *Zuelania guidonia* alcanzan alturas por arriba de los 45 m. El área basal por hectárea mostró diferencias que evidencian mayor tamaño en los troncos y en este sentido *Clethra microphylla*, *V. guatemalensis*, *Zuelania guidonia* y *Pouteria torta* son las especies de mayor contribución de área basal del tallo, en contraste, *Calatola laevigata* e *I. quaternata* son las especies que contribuyen con menor área basal por hectárea. Para la cobertura de la copa las especies *V. guatemalensis*, *Pseudolmedia oxyphyllaria*, *Zuelania guidonia*, *Ocotea dendrodaphne* y *Clethra macrophylla* que son las especies que aportan mayor cobertura de sombra. Es notorio que las especies de árboles que presentan mayor área basal por hectárea son las que tienen también más cobertura. La distribución de frecuencias de la altura de las diecinueve especies de árboles de la selva tiene amplia variabilidad al compararse interespecíficamente. Se observa que las diecinueve especies convergen en tallas de altura de 16.5 a 22 m, ya que son las categorías más representadas porcentualmente, así, las únicas especies representadas en el mayor número de categorías son *Virola guatemalensis* y *Roupala montana* en diez de once categorías para la altura. Para el diámetro a la altura del pecho las diecinueve especies de leñosas convergen en la categoría de 40 cm ya que es la categoría mas representada y las especies *Pouteria torta* y *V. guatemalensis* son las especies representadas en el mayor número de categorías con seis. Para el caso del parámetro que se refiere a la cobertura de la copa todos los árboles convergen en los 100 m², siendo únicamente las especies *Clethra macrophylla* y *V. guatemalensis* las especies con el mayor número de categorías representadas con seis. La comparación de las alturas promedio de las diecinueve especies de árboles mostró diferencias significativas ($F_{(18,754)} = 12, P < 0.001$). Fue así que *Turpinia occidentalis* (14.96 m), *Calatola laevigata* (15.38 m), *P. chiapensis* (15.84 m) y *Pouteria lucentifolia* (15.92 m) fueron las leñosas con menor altura promedio y en contraste *Zuelania guidonia* (28.24 m) fue en promedio la especie con mayor altura antecedida por *Clethra macrophylla* y *V. guatemalensis*. El diámetro a la altura del pecho también mostró diferencias significativas entre las especies de árboles ($F_{(18,724)} = 10, P < 0.001$). En este caso las especies *Clethra macrophylla* (49.77 cm) y *Zuelania guidonia* (47.75 cm) fueron las leñosas con mayor diámetro promedio. Asimismo la especie *Virola guatemalensis* (39.71 cm) presenta un

valor promedio del diámetro entre el grupo de las leñosas de mayor tamaño y un grupo intermedio conformado por once especies y destaca como la especie con menor diámetro promedio *Calatola laeviegata* (24.66 cm). Para el caso de la cobertura de la copa también se obtuvieron diferencias significativas ($F_{(18,757)}=6, P < 0.001$) aunque en la comparación de los valores promedio entre las especies de árboles fueron menos contrastantes. Las especies *Zuelania guidonia* (103.51 m²) y *Clethra macrophylla* (82.22 m²) fueron las especie con mayor cobertura de copa promedio y *Calatola laeviegata* (25.46 m²) la de menor cobertura. En contraste las especies *Ocotea dendrodaphne*, *P.torta* y *Virola guatemalensis* fueron similares entre si, de la misma forma *Roupala montana* e *Ilex discolor* tiene una similitud en cuanto a su promedio y el resto de las especies de leñosas conforman un grupo que no pasa de los 43 m². Los valores de A_{max} y C_{max} se relacionaron y se obtuvo una asociación positiva ($F=35; P < 0.0001$), lo cual indica que como los valores más altos correspondieron a las especies *Clethra macrophylla*, *Zuelania guidonia* y *Virola guatemalensis*, árboles que corresponderían al estrato más alto del dosel; se puede observar que los árboles del estrato medio tienen valores menores a los 48 m de altura y por debajo de los 350 m² de cobertura, *Calatola laeviegata* de acuerdo a esta función alométrica es la que presenta el más bajo coeficiente (Fig.1). El modelo alométrico logarítmico de las diecinueve especies mostró distintos patrones en las relaciones de tamaño. De las diecinueve proyecciones lineales sólo *I. quaternata* y *Turpinia occidentalis* no presentaron asociación entre el diámetro a la altura del pecho vs. altura de los árboles. Por el contrario, las especies *Ocotea dendrodaphne*, *Heliocarpus appendiculatus* y *Clethra macrophylla* presentaron los mayores coeficientes de correlación (Fig. 2). Por otro lado, las relaciones alométricas entre el diámetro a la altura del pecho vs. cobertura de la copa también mostraron variación interespecífica. Destacan las asociaciones lineales entre los dos atributos de crecimiento de las especies *Zuelania guidonia*, *Virola guatemalensis* y *Ocotea dendrodaphne* que son las especies con mayor coeficiente de correlación. Así, las especies *Guarea glabra*, *Dendropanax arboreus* y *Turpinia occidentalis* no presentaron asociación entre el diámetro a la altura del pecho vs. cobertura de la copa (Fig. 3).

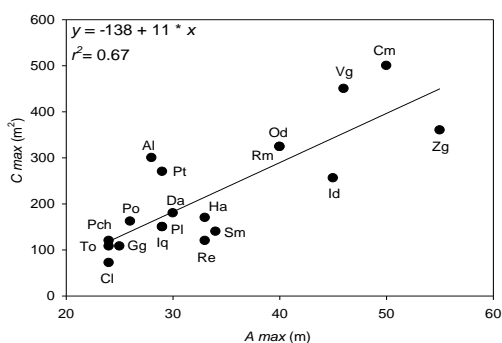


Fig 1. Relación entre la A_{max} y C_{max} de las 19 especies de selva alta perennifolia.

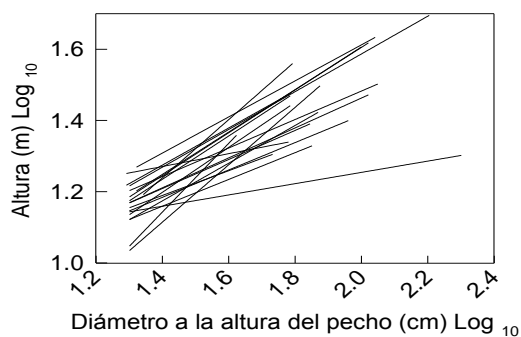


Fig.2. Asociaciones lineales del DAP vs. ALT.

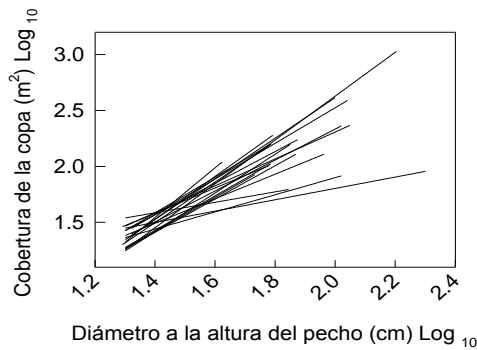


Fig. 3. Asociaciones lineales del DAP vs. COB.

Conclusión

Las distribuciones de frecuencias de los tres atributos de tamaño difieren entre especies. Además, la leñosa *V. guatemalensis* tiene porcentualmente más individuos representados en las distintas categorías de tamaño. La dasometría estimada para las diecinueve especies de árboles presenta amplia variabilidad. De ahí que la densidad de individuos es mayor para las especies *Pseudolmedia oxyphyllaria* y *R. edulis*. En cambio, las especies que aportan mayor área basal por hectárea fueron *Clethra macrophylla*, *V. guatemalensis* y *Zuelania guidonia*. Siendo estas dos últimas las que cubren mayor superficie con su cobertura de copa. En cuanto a los valores de altura de los árboles, se diferencian por su mayor A_{max} la leñosa *Z. guidonia*, *C. macrophylla* y *V. guatemalensis*. Este patrón coincide para el caso de la C_{max} , puesto que las especies de mayor A_{max} fueron las que tienen mayor cobertura. La asociación positiva entre la A_{max} y C_{max} , indica que existe un costo-beneficio en cuanto a la inversión de cada especie al crecimiento vertical y horizontal dentro del dosel. La relación alométrica entre el diámetro a la altura del pecho vs. altura del árbol difiere entre grupos de especies, debido a que el más alto coeficiente alométrico lo presento *R. edulis*. Además, la asociación entre el diámetro del tronco vs. cobertura de la copa, también difiere interespecíficamente, patrones que se definen en cuanto a su posición dentro del dosel de la selva alta perennifolia.

REFERENCIAS

- Aiba SI, Kohyama T. 1996. Tree species stratification in relation to allometry and demography in a warm-temperate rain forest. *Journal of Ecology* 84: 207-218.
- Akashi N. 2006. Height growth of young larch (*Larix kaempferi*) in relation to the frequency of deer browsing damage in Hokkaido, Japan. *Journal Forest Research* 11: 153-156.
- Archibald S, Bond WJ. 2003. Growing tall vs growing wide: tree architecture and allometry of *Acacia karroo* in forest, savanna and arid environments. *Oikos* 102: 3-14.
- [Gayon J. 2000. History of the concept of allometry. *American Zoologist* 40: 748-758.](#)
- King DA. 1990. Allometry of saplings and understory trees of a panamanian forest. *Functional Ecology* 4: 27-32.
- Kohyama T. 1987. Significance of architecture and allometry in saplings. *Functional Ecology* 1: 399-404.
- Kohyama T, Hotta M. 1990. Significance of allometry in tropical saplings. *Functional Ecology* 4: 515-521.

- Kohyama T. 1992. Size-structured multi-species model of rain forest trees. *Functional Ecology* 6: 206-212.
- Kohyama T. 1993. Size-structure tree population in gap-dynamic forest: the forest architecture hypothesis for the stable co-existence of species. *Journal of Ecology* 81: 131-143.
- Kohyama T, Suzuki E, Partomihardjo T, Yamada T, Kubo T. 2003. Tree species differentiation in growth, recruitment and allometry in relation to maximum height in a Bornean mixed dipterocarp forest. *Journal of Ecology* 91: 797-806.
- Küppers M. 1989. Ecological significance of above ground architectural patterns in Woody plants: a question of cost-benefit relationships. *Trends in Evolution and Ecology* 4: 375-378.
- López-Serrano FR, García-Morote A, Andrés-Abellán M, Tendero A, Del Cerro A. 2005. Site and weather effects in allometries: a simple approach to climate change effect on pines. *Forest Ecology and Management* 215: 251-270.
- Martínez AJ, López-Portillo J. 2003. Allometry of *Prosopis glandulosa* var. *torreyana* along a topographic gradient in the Chihuahuan desert. *Journal of Vegetation Science* 14: 111-120.
- Méndez-Alonzo R, López-Portillo J, Rivera-Monroy VH. 2008. Latitudinal variation in leaf and tree traits of the mangrove *Avicennia germinans* (Avicenniaceae) in the Central Region of the Gulf of Mexico. *Biotropica* 4: 449-456.
- Niklas KJ. 1994. Plant allometry: the scaling of form and process. Universidad de Chicago. 395 pp.
- Poorter L, Bongers F, Sterck FJ, Wöll H. 2005. Beyond the regeneration phase: differentiation of height light trajectories among tropical tree species. *Journal of Ecology* 93: 256-267.
- Poorter L, Bongers L, Bongers F. 2006. Architecture of 54 moist forest tree species: traits, trade-offs, and functional groups. *Ecology* 87: 1289-1301.
- Poorter L, Hawthorne W, Bongers F, Sheil D. 2008. Maximum size distributions in tropical forest communities: relationships with rainfall and disturbance. *Journal of Ecology* 96: 495-504.
- Richards PW. 1996. The tropical rain forest. Second edition. Cambridge University Press, Cambridge.
- Sterck FJ, Bongers F. 1998. Ontogenetic changes in size, allometry and mechanical design of tropical rain forest trees. *American Journal of Botany* 85: 266-272.
- Takahashi K, Kohyama T. 1997. Crow architecture of two understory palm species of the genus *Licuala* in a tropical rain forest. *Plant Species Biology* 12: 35-41.
- Takahashi K, Seino T, Kohyama T. 2001. Responses to canopy openings in architectural development of saplings in eight deciduous broad-leaved tree species. *Canadian Journal of Forest Research* 31: 1336-1347.
- Turner IM, Gong WK, Ong JE, Bujang JS, Kohyama T. 1995. The architecture and allometry of mangrove saplings. *Functional Ecology* 9: 205-212.

Comparación de Estimadores de Regresión del Total en Poblaciones Estratificadas y Complejas Bajo Tres Propuestas de la Matriz de Varianzas

Flaviano Godínez Jaimes¹, Ignacio Méndez Ramírez², Ma. Natividad Nava Hernández³

¹Universidad Autónoma de Guerrero. fgodinezj@gmail.com

²IIMAS. UNAM. imendez@servidor.unam.mx

³Universidad Autónoma de Guerrero

Resumen

Los estimadores de regresión del total consideran que la variable de interés, Y , y un conjunto de variables auxiliares X_1, X_2, \dots, X_p están relacionadas mediante un modelo de regresión lineal, esto es, suponen que $E[\mathbf{y}] = \mathbf{X}\beta$, $V[\mathbf{y}] = \mathbf{V}$. Los estimadores que consideran el modelo anterior son Basado en Modelo, Asistido por Modelo y Cosméricamente Calibrado. Generalmente la matriz de varianzas y covarianzas \mathbf{V} es desconocida en muestreo de conglomerados y en muestras complejas que involucren conglomerados. Una solución es usar $\mathbf{V}=\mathbf{I}$, $\mathbf{\Pi}$ o \mathbf{G} donde $\mathbf{\Pi}$ depende de las probabilidades de inclusión y \mathbf{G} es una estimación de \mathbf{V} mediante las ecuaciones de estimación generalizadas. Mediante simulación se hace una comparación de los estimadores de regresión del total Basado en Modelo, Asistido por Modelo y Cosméricamente Calibrado en poblaciones complejas considerando las tres propuestas de la matriz \mathbf{V} y los diseños estratificado y Complejo

Palabras clave: GREG, muestras complejas, estimador basado en modelo, ecuaciones de estimación generalizada.

Introducción

Actualmente, muchas encuestas nacionales son estratificadas y con conglomerados, es decir, la población es compleja. Los diseños de muestreo que toman en cuenta esta característica de la población deben ser estratificados y multietápicos y los estimadores deben incorporar esta información.

Este trabajo se centra en la estimación del total de la variable de interés y que toma los valores $\{y_1, \dots, y_M\}$ en una población finita de M unidades $P = \{1, \dots, M\}$:

$$T = \sum_{i \in U} y_i = \sum_{i=1}^N y_i$$

Una **población es estratificada** si se puede particionar en L estratos donde las unidades dentro de un estrato son muy homogéneas y muy heterogéneas entre estratos. Si M^h es el tamaño del estrato h , $h=1, \dots, L$, entonces el tamaño de P es $M = \sum_{h=1}^L M^h$. EL muestreo

estratificado se realiza tomando una muestra aleatoria independiente en cada estrato. Una **población es conglomerada** si las unidades están agrupadas de manera natural en N^h conglomerados o unidades primarias de muestreo (UPM) cada una con N_i^h elementos o unidades secundarias de muestreo (USM) $i=1, \dots, N^h$. Una muestra por conglomerados en

dos etapas o muestreo bietápico consiste en seleccionar una muestra de UPM y dentro de las UPM seleccionadas tomar una muestra aleatoria de USM. Una **población compleja** se forma por L estratos cada estrato con N^h UPM con N_i^h USM, $h = 1, \dots, L$; $i = 1, \dots, N^h$.

Por tanto, los tamaños de estrato y de la población son $M^h = \sum_{i=1}^{N^h} N_i^h$ y $M = \sum_{h=1}^L M^h$.

Un esquema de muestreo para una población compleja es en cada estrato tomar una muestra aleatoria simple sin reemplazo (MAS) de UPM y en las UPM seleccionadas usar cualquier forma de muestreo incluso diferente. En particular, se puede elegir con MAS n^h de las N^h UPM y con MAS n_i^h de las N_i^h USM, por lo que el vector de probabilidades de inclusión en el estrato h , de orden $n^h n_i^h$, es $\boldsymbol{\pi}^h = (\pi_i)^h = (\pi_{h_i} \pi_{k|i})^h = \left(\frac{n^h}{N^h} \frac{n_i^h}{N_i^h} \dots \frac{n^h}{N^h} \frac{n_i^h}{N_i^h} \right)^T$ y en la muestra completa, de orden $L n^h n_i^h$, es $\boldsymbol{\pi} = (\boldsymbol{\pi}^1, \boldsymbol{\pi}^2, \dots, \boldsymbol{\pi}^L)^T$.

Los estimadores de regresión del total (ERT) son útiles cuando hay información de variables auxiliares para todos los elementos de la población y las variables auxiliares están asociadas linealmente con la variable de estudio. Es decir, están relacionadas mediante el modelo de regresión lineal

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i; \quad i = 1, \dots, N$$

donde $\boldsymbol{\beta}$ es el vector de parámetros de regresión, y_i y $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ son la observación i -ésima de la variable de interés y de las variables auxiliares X_1, X_2, \dots, X_p además se supone que los errores, e_i , tienen media cero, varianza σ_i^2 y covarianza σ_{ij} . Matricialmente el MRL se puede escribir como $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ con $E(\boldsymbol{\varepsilon}) = 0$ y $V(\boldsymbol{\varepsilon}) = \mathbf{V}$ y por tanto $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ y $V(\mathbf{Y}) = \mathbf{V}$.

Estimador de Regresión del Total Basado en Modelo

El predictor óptimo para $\boldsymbol{\beta}$ basado en modelo es $\hat{\boldsymbol{\beta}} = (\mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{Y}_s$, donde $\mathbf{X}_s, \mathbf{Y}_s$ y \mathbf{V}_s son la muestra de \mathbf{X}, \mathbf{Y} y \mathbf{V} . El ERT basado en modelo es

$$\hat{T}_{BM} = \mathbf{1}_s^T \mathbf{Y}_s + \mathbf{1}_r^T \mathbf{X}_r \hat{\boldsymbol{\beta}} = \mathbf{g}_s^T \mathbf{Y}_s,$$

donde $\mathbf{g}_s^T = \mathbf{1}_s^T + \mathbf{a}_s^T$ es el vector de pesos y $\mathbf{a}_s = \mathbf{V}_s^{-1} \mathbf{X}_s (\mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_r^T \mathbf{1}_r$. Dos estimadores de la varianza aproximada de \hat{T}_{BM} , son $V_1(\hat{T}_{BM}) = \sum_s \mathbf{g}_i^T (\mathbf{r}_i \mathbf{r}_i^T) \mathbf{g}_i$ y $V_2(\hat{T}_{BM}) = \sum_s \mathbf{a}_i^T (\mathbf{r}_i \mathbf{r}_i^T) \mathbf{a}_i$, donde $\mathbf{g}_i, \mathbf{r}_i$ y \mathbf{a}_i son la parte correspondiente al cluster i en la muestra de $\mathbf{g}_s, \mathbf{r}_s = \mathbf{Y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}$. La expresión para \hat{T}_{BM} es válida para cualquier tipo de muestreo pero $V_1(\hat{T}_{BM})$ y $V_2(\hat{T}_{BM})$ son específicos cuando la población es por conglomerados.

Estimador de Regresión del Total Asistido por Modelo

En el enfoque AM y bajo el modelo de trabajo $E_M(\mathbf{Y}) = \mathbf{X}\beta$, $V_M(\mathbf{Y}) = \mathbf{V} = \text{diag}(\sigma_1^2 \dots \sigma_N^2)$, el ERT es:

$$\hat{T}_{AM} = \mathbf{1}^T \Pi_s^{-1} \mathbf{Y}_s + (\mathbf{1}^T \mathbf{X} - \mathbf{1}^T \mathbf{V}_s^{-1} \Pi_s^{-1} \mathbf{X}_s) \hat{\mathbf{B}};$$

donde $\hat{\mathbf{B}} = \mathbf{A}^{-1} \mathbf{X}_s^T \mathbf{V}_s^{-1} \Pi_s^{-1} \mathbf{Y}_s$, $\mathbf{A}_{ss} = \mathbf{X}_s^T \mathbf{V}_s^{-1} \Pi_s^{-1} \mathbf{X}_s$, $g_{isB} = 1 + (\mathbf{1}^T \mathbf{X} - \mathbf{1}^T \mathbf{V}_s^{-1} \Pi_s^{-1} \mathbf{X}_s)^T \mathbf{A}_{ss}^{-1} \mathbf{x}_i$, y $\Pi_s = \text{diag}(\pi_i)$.

\hat{T}_{AM} es aproximadamente insesgado en diseño, si el número de clusters es grande, e insesgado bajo el modelo de trabajo.

La varianza aproximada estimada es

$$\hat{V}(\hat{T}_{AM}) = \sum_{i=1}^{n^h} \sum_{j=1}^{n_i^h} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \right) \frac{\hat{t}_{Ei}}{\pi_i} \frac{\hat{t}_{Ej}}{\pi_j} - \sum_{i=1}^{n^h} \frac{1}{\pi_i} \left(\frac{1}{\pi_i} - 1 \right) \hat{V}_{BEi} + \sum_{i=1}^{n^h} \frac{\hat{V}_{BEi}}{\pi_i^2}$$

$$\text{con } \hat{V}_{BEi} = \sum_{k=1}^{n^h} \sum_{l=1}^{n_i^h} \left(\frac{\pi_{kl|i} - \pi_{k|i} \pi_{l|i}}{\pi_{kl|i}} \right) \frac{g_{kSB} e_{ks}}{\pi_{k|i}} \frac{g_{lSB} e_{ls}}{\pi_{l|i}}, \quad \hat{t}_{Ei} = \sum_{k=1}^{n_i^h} g_{kSB} e_{ks} / \pi_{k|i} \quad \text{y} \quad e_{ks} = y_k - \mathbf{x}_k' \hat{\mathbf{B}}.$$

Ambos, \hat{T}_{BM} y \hat{T}_{AM} son válidos para cualquier tipo de muestreo pero los estimadores $V_1(\hat{T}_{BM})$, $V_2(\hat{T}_{BM})$ y $\hat{V}(\hat{T}_{AM})$ son específicos cuando la población es por conglomerados si la población es compleja los ERT y sus varianzas estimadas se pueden obtener:

a) Bajo el diseño estratificado $\hat{T}_E = \sum_{l=1}^L \hat{T}_*^h$ y $\hat{V}(\hat{T}_E) = \sum_{l=1}^L \hat{V}(\hat{T}_*^h)$ donde $\hat{T}_*^h = \hat{T}_{BM}$ o \hat{T}_{AM} y $\hat{V}(\hat{T}_*^h) = \hat{V}(\hat{T}_{BM})$ o $\hat{V}(\hat{T}_{AM})$ son calculados en el estrato h usando el vector de probabilidades de inclusión π^h .

b) Bajo el diseño complejo $\hat{T}_C = \hat{T}_{BM}$ o \hat{T}_{AM} y $\hat{V}(\hat{T}_C) = \hat{V}(\hat{T}_{BM})$ o $\hat{V}(\hat{T}_{AM})$ calculados usando el vector de probabilidades de inclusión π .

Simulación

Los datos en el estrato h se generan de acuerdo con un modelo de efectos aleatorios

$$Y_{ij}^h = \mu^h + A_i^h + E_{ij}^h = \beta_1^h + \beta_1^h X_{1ij}^h + \beta_2^h X_{2ij}^h + \dots + \beta_p^h X_{pij}^h + A_i^h + E_{ij}^h$$

$h=1, \dots, 10$; $i=1, \dots, N^h$; $j=1, \dots, N_i^h$ donde $A_i^h \sim N(0, \sigma_A^2)$ y $E_{ij}^h \sim N(0, \sigma_E^2)$ son independientes; $\beta = (5,400k_b, 1000k_b, 3000k_b)^T$ fijo y $X_k^h \sim N(\mu_k^h, \sigma_k^2)$, $k=1, 2, 3$. con $(\mu_1^h, \mu_2^h, \mu_3^h) = (1000 + 100000h, 5000 + 100000h, 9000 + 100000h)$, $\sigma_k^2 = 400^2, 1000^2, 30000^2$. Por tanto $E_M(\mathbf{Y}) = \mathbf{X}\beta$ y $V_M(\mathbf{Y}) = \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_{N^h})$ con $\mathbf{V}_i = \sigma^2 \mathbf{R}_i$. \mathbf{R}_i es la matriz de correlaciones intercambiable que fuera de la diagonal tiene a $\rho = \sigma_A^2 / \sigma^2 = \sigma_A^2 / (\sigma_A^2 + \sigma_E^2)$, el coeficiente de correlación intraconglomerado cuantifica la variación entre los clusters. Es conocido que

el muestreo por conglomerados es menos eficiente que el muestreo aleatorio simple pero no se puede afirmar lo mismo en muestras complejas.

El grado de asociación lineal entre la variable de interés y las variables auxiliares puede afectar el desempeño de los ERT (Dever y Valliant, 2006). Una medida de la asociación es $R^2 = SCR/SCT$, donde $SCR = \beta^T \mathbf{X}^T \mathbf{y} - N^{-1}(\mathbf{1}^T \mathbf{y})^2$ y $SCT = \mathbf{y}^T \mathbf{y} - N^{-1}(\mathbf{1}^T \mathbf{y})^2$. Para obtener semejanzas débil y fuerte dentro de los conglomerados se tomaron $\rho = 0.33$ y 0.95 que se obtuvieron dando a σ_A^2 y σ_E^2 valores de 25000, 5000 y 25000, 1000 respectivamente. Mientras que la asociación débil y fuerte entre la variable respuesta y las variables auxiliares, $R^2 = 0.33$ y 0.95 , se obtuvo con valores de k_b de 0.0001, 0.00002, 0.00005 y 0.00001.

En los estimadores \mathbf{V}_s se sustituye por \mathbf{I}_s , $\mathbf{\Pi}_s$ y \mathbf{G}_s , es decir, considerando la matriz identidad, una matriz que depende de las probabilidades de inclusión y una matriz que se obtiene usando ecuaciones de estimación generalizadas generada con:

1. Estimar β (con $\mathbf{V}_s = \mathbf{I}_s$) con mínimos cuadrados ordinarios y llamarlo \mathbf{b}_0 .
2. Usar \mathbf{b}_0 para calcular los residuos, $e_i^h = y_i^h - \mathbf{x}_i^h \mathbf{b}_0$, las correlaciones

$$\hat{\rho} = \frac{1}{K_1 - p} \sum_{i=1}^{n^h} \sum_{j \leq n_i=1}^{n^h} e_{ij} e_{i,j+1}, \quad \hat{\sigma}^2 = tr \left[\frac{\sum_{j=1}^{n^h} (\mathbf{Y}_s - \mathbf{X}_s \mathbf{b}) (\mathbf{Y}_s - \mathbf{X}_s \mathbf{b})^T}{N - p} \right] \text{ y determinar } \hat{\mathbf{R}} \text{ y } \hat{\mathbf{G}}.$$

3. Calcular $\mathbf{b} = (\mathbf{X}_s^T \hat{\mathbf{R}}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \hat{\mathbf{R}}^{-1} \mathbf{Y}_s$.
4. Regresar al paso 2 e iterar el proceso hasta lograr la convergencia.

Se estudian cuatro escenarios según los valores de ρ y R^2 : E1) 0.33-0.33, E2) 0.33-0.95, E3) 0.95-0.33, E4) 0.95-0.95. Se obtuvieron 15000 muestras de tamaño 200 con el esquema de muestreo mencionado y se calcularon los dos ERT bajo las tres propuestas de varianza y los dos diseños. Los estimadores estudiados se comparan utilizando a) Sesgo Relativo: $\hat{E}((\hat{T} - T)/T)$, b) Error Cuadrático Medio Relativo: $\hat{E}((\hat{T} - T)^2)/T$ y c) Cubrimiento de intervalos del 95% de confianza: $\hat{T} \pm 2\hat{v}(\hat{T})$.

Resultados

Se usa BMI, BMP y BMG para indicar que en el enfoque BM, \mathbf{V}_s se sustituye por \mathbf{I}_s , $\mathbf{\Pi}_s$ y \mathbf{G}_s respectivamente y de manera similar para AM y CC. En el Cuadro 1, se observa que cuando hay mas asociación lineal entre la variable de interés y las variables auxiliares ($R^2 = 0.95$) los estimadores estudiados tienen menor SR y ECMR. Es mas frecuente que sobrestimen al total poblacional especialmente en el diseño estratificado con $R^2 = 0.95$ y lo contrario con $R^2 = 0.33$. Los estimadores generalmente tienen mayor ECMR y mayor dispersión bajo el diseño estratificado que en el diseño complejo.

Solo veintidós de cuatrocientos cincuenta y seis combinaciones de estimadores con sus varianzas tuvieron cuando mucho una desviación de 0.03 del valor nominal 0.95. De estos, dieciocho estimadores usan $\mathbf{V}_s = \mathbf{I}_s$ y cuatro $\mathbf{V}_s = \mathbf{\Pi}_s$. En relación al enfoque, catorce estimadores usan el BM y ocho el AM. En cuanto al diseño, trece son estratificados y nueve complejos.

Cuadro 1. Sesgo y Error Cuadrático Medio Relativos de los estimadores estudiados.

| | | 33 95 | | 95 95 | | 95 33 | | 33 33 | |
|-------|-----|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|-------|
| ESTIM | | SR [§] | ECMR | SR [§] | ECMR | SR [§] | ECMR | SR [§] | ECMR |
| ESTR | BMI | 507 | 12.03 | 3857 | 13.36 | -11793 | 76.44 | -10961 | 54.04 |
| | CCI | 7 | 11.88 | 485 | 13.01 | -2797 | 68.56 | -3799 | 55.65 |
| | AMI | 6 | 11.87 | 491 | 13.00 | -2792 | 68.54 | -3811 | 55.63 |
| | BMP | 25 | 11.86 | 647 | 12.97 | -3147 | 68.65 | -4070 | 55.41 |
| | CCP | 7 | 11.88 | 485 | 13.01 | -2797 | 68.56 | -3799 | 55.65 |
| | AMP | -7 | 12.02 | 378 | 13.06 | -2918 | 68.75 | -3715 | 56.25 |
| | BMG | 550 | 11.96 | 4184 | 13.16 | -12535 | 74.21 | -10823 | 53.90 |
| | CCG | 5 | 11.80 | 478 | 12.70 | -2589 | 65.86 | -3403 | 55.83 |
| | AMG | 3 | 11.79 | 487 | 12.71 | -2562 | 65.84 | -3417 | 55.81 |
| | BMI | 248 | 10.79 | 2918 | 12.35 | -16002 | 71.14 | -5519 | 47.67 |
| | CCI | 50 | 10.98 | 8 | 13.03 | -699 | 65.43 | 83 | 53.28 |
| | AMI | 35 | 10.97 | -51 | 12.95 | -836 | 65.53 | -29 | 53.02 |
| COMP | BMP | 731 | 10.84 | 652 | 11.87 | -8232 | 65.13 | 329 | 50.69 |
| | CCP | 50 | 10.98 | 8 | 13.03 | -699 | 65.43 | 83 | 53.28 |
| | AMP | 15 | 11.02 | -103 | 12.96 | -847 | 65.65 | 29 | 53.18 |
| | BMG | 664 | 10.87 | 2678 | 10.23 | -37379 | 83.94 | -5456 | 46.93 |
| | CCG | 46 | 10.96 | -111 | 12.86 | -1245 | 64.60 | 80 | 53.20 |
| | AMG | 29 | 10.95 | -105 | 12.95 | -779 | 64.99 | -11 | 53.02 |

§ Los valores son múltiplos de 10^8

Conclusiones

Los estimadores estudiados tienen menor SR, ECMR y sobrestiman el Total cuando hay más asociación lineal entre la variable de interés y las variables auxiliares, $R^2=0.95$. En términos de cubrimiento, los estimadores BM fueron mejores que AM. La mejor propuesta de varianza es $V=I$ seguida muy lejos de $V=II$ mientras que $V=G$ que no cumplió con el valor nominal. Se obtuvieron aproximadamente iguales resultados con ambos diseños Estratificado y Complejo. BM fue mucho mejor cuando $\rho=0.33$, es decir cuando la población es aleatoria.

REFERENCIAS

- Brewer, K. (2002). Combined Survey Sampling Inference, Weighing Basu's Elephants. Arnold: London.
- Dever, J. A. y Valliant, R. (2006). A comparison of Model-Based and Model-Assisted Estimators under Ignorable and Non-Ignorable Nonresponse. Proceedings of the Survey Research Methods Section. American Statistical Association.
- Godínez, J. F. y Méndez, R. I. (2008). Desempeño en muestras complejas de tres estimadores de regresión del total. Memorias del XXII Foro Nacional de Estadística. INEGI: Mexico.
- Raj, D. (1968). Sampling Theory. McGraw-Hill: New York.
- Särndal, C.E., Swensson, B. y Wretman, J. (1992). Model Assisted Survey Sampling. Springer Verlag: New York.
- Valliant, R., Dorfman, A.H. y Royall, R.M. (2000). Finite Population Sampling and Inference: A Prediction Approach. John Wiley & Sons.: New York.

Un Estudio de la Evolución de Epidemias de Influenza en Estados Unidos y la Posible Asociación con el Factor Climatológico

Agustin Jaime Garcia Banda¹, Luis Cruz-Kuri², Ismael Sosa Galindo³

¹Facultad de Ciencias Administrativas y Sociales. U.V. jaimegarciabanda@yahoo.com

²Instituto de Ciencias Basicas. U.V. kruz1111@yahoo.com.mx

³Facultad de Ciencias Administrativas y Sociales. U.V. isoga77@yahoo.com.mx

1. Introducción

Los recientes eventos de la aparición de un nuevo tipo de influenza, ahora denominada A(H1N1), atrajeron la atención a nivel mundial de los especialistas de disciplinas afines, así como del público en general. Para nuestro reporte, con información recabada del CDC (Centers for Disease Control and Prevention) de Atlanta y la organización WHO (World Health Organization), se consiguieron datos estadísticos de la evolución de esta epidemia en los Estados Unidos, al igual que en otros países. Adicionalmente, se recabaron datos de las series observadas de tiempo correspondientes a temperaturas mínima, media y máxima, en distintas ciudades de ese país. Si se examinan las gráficas correspondientes de la evolución de la influenza típica (para la cual se tiene información de varios años) se nota un comportamiento cíclico, donde el número de casos aumenta durante el invierno y disminuye durante las estaciones más benignas del año (primavera y verano). Por todo esto, realizamos un análisis de posibles vinculaciones entre las series epidemiológicas y las series climatológicas. El estudio de las posibles asociaciones se realizó a nivel estatal para algunos estados de ese país, seleccionando series de tiempo de temperatura de ciudades representativas de cada estado considerado; estas últimas se pueden recabar, por ejemplo, en NOAA (National Oceanic and Atmospheric Administration) o en NCDC (National Climatic Data Center). A manera de ilustración, algunos estados seleccionados fueron: Texas, Wisconsin, Illinois, California y Arizona. Por razones de espacio, para el presente trabajo, se reportan nuestros hallazgos únicamente como ejemplo de lo que puede hacerse para todas las entidades del país. Para la distribución espacial de la epidemia a nivel de todo Estados Unidos se hace un seguimiento de forma gráfica con una serie de tiempo de mapas, con desglosamiento a nivel estatal, a lo largo de trece semanas del 2008 y aproximadamente veinte semanas del 2009.

2. Reportes Semanales de Vigilancia de Influenza del CDC

Para todos los tipos de influenza, la división a cargo, desde hace muchos años elabora reportes semanales de la distribución de esta enfermedad en todos los estados de la unión americana y en particular presenta mapas codificados por color según sea la presencia de casos (desde ligera, con difusión esporádica, difusión regional, hasta difusión generalizada). A manera de ilustración, el color blanco corresponde a ausencia de información, en tanto que el color café en tono oscuro corresponde a una difusión generalizada. A continuación se presentan cuatro mapas correspondientes, respectivamente, a la cuarta semana de enero, la segunda semana de febrero, la primera semana de marzo y la cuarta semana de marzo, todas del 2009. Aunque se pueden presentar los mapas de todas las semanas de varios años, por consideraciones de espacio no se hace esto, y las que aquí se presentan se seleccionaron para ilustrar distintos patrones de presencia de la enfermedad. Por ejemplo, en la Fig.1,

únicamente los estados de Texas, Virginia y Nueva Jersey tenían una difusión generalizada de influenza en esa semana de invierno crudo de 2009. En lo correspondiente a la Fig. 2, para la segunda semana de febrero, el número de estados con presencia generalizada de influenza había aumentado y ahora abarcaba, aparte de los anteriores, los estados de Nevada, Arizona, Colorado, Georgia, Carolina del Norte y Alabama, entre otros. Para la primera semana de marzo de 2009, el número de estados con presencia generalizada de influenza se había incrementado y ahora casi todos los estados del norte de ese país quedaban incluidos exceptuando a Minnesota y Wisconsin (ver Fig. 3). Finalmente, para la cuarta semana de marzo, se redujo el número de estados con presencia generalizada, persistiendo en Washington, Idaho, Oregon y Colorado y además apareciendo por primera vez California. Una inspección de este tipo de mapas a largo de varios años, refleja un patrón de comportamiento periódico, que de alguna manera está relacionado con la componente climatológica.

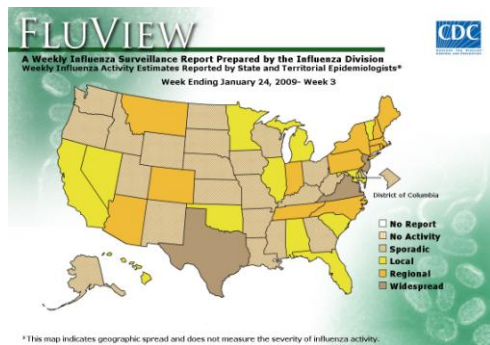


Fig. 1. Semana del 24 de enero de 2009.
Fuente: CDC (Centers for Disease Control and Prevention).

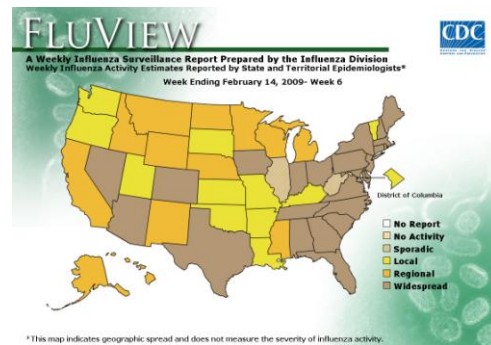


Fig. 2. Semana del 14 de febrero de 2009.
Fuente: CDC (Centers for Disease Control and Prevention).

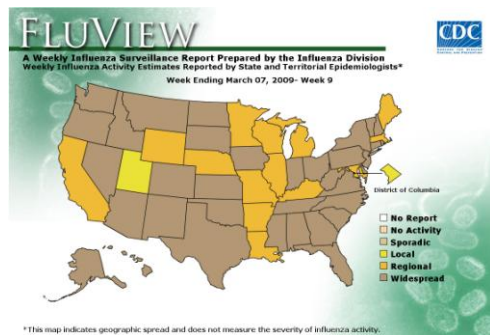


Fig. 3. Semana del 7 de marzo de 2009.
Fuente: CDC (Centers for Disease Control and Prevention).

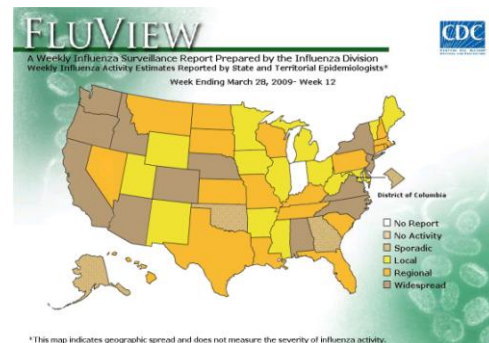


Fig. 4. Semana del 28 de marzo de 2009.
Fuente: CDC (Centers for Disease Control and Prevention).

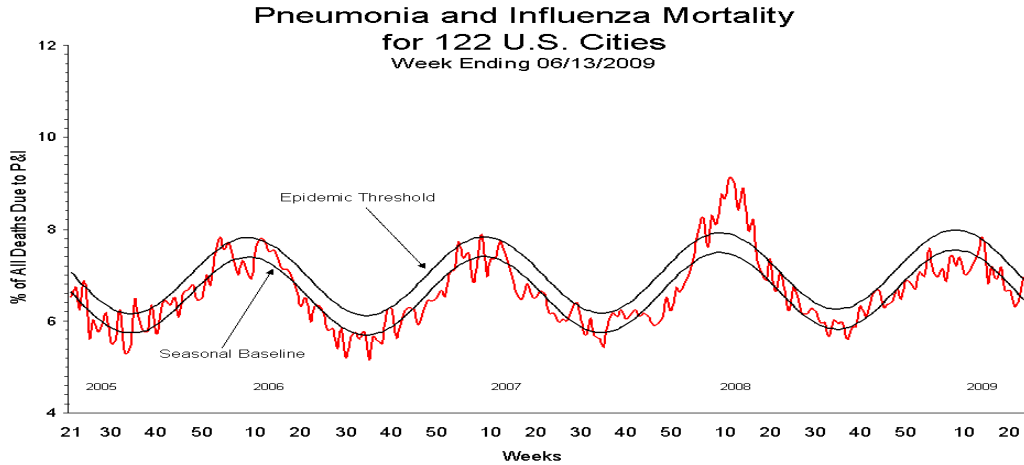


Fig. 5. Mortalidad atribuida a influenza y neumonía. Total de casos agregados semanalmente en 122 ciudades de Estados Unidos. FUENTE: CDC.

3. Análisis Estadístico

Por consideraciones de espacio, en el presente reporte solo se analizan los datos recopilados para dos estados de la nación norteamericana, a saber, Texas y Wisconsin. Tal como ya se mencionó en la introducción, para cada estado se seleccionaron cinco ciudades representativas; las cuales son: Houston, Dallas, Amarillo, Brownsville y El Paso, todas estas en el estado de Texas. Para el estado de Wisconsin, las ciudades seleccionadas fueron: Madison, Superior, Green Bay, Milwaukee y Wausau. Para cada ciudad, se registraron las temperaturas mínima, media y máxima, siendo la frecuencia de registros semanal. Para cada uno de los tres tipos de temperatura, se obtuvo una temperatura representativa de todo el estado mediante el cálculo de promedios simples. Por otra parte, se recopiló la información epidemiológica con frecuencia semanal y con cobertura estatal. Las variables epidemiológicas consideradas fueron: Porcentaje y número total de muertes (todas las causas).

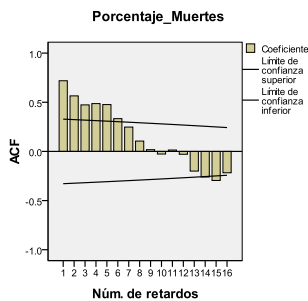


Fig.6. Autocorrelaciones desde 0 hasta 16 semanas de desplazamiento para la variable “Porcentaje de Muertes”.

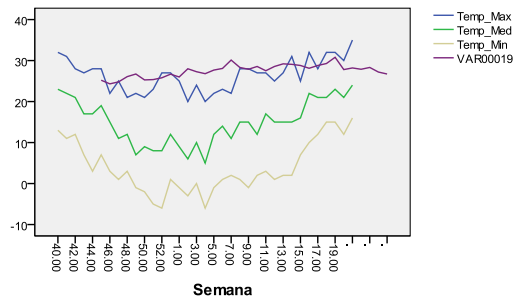


Fig.7. Series de tiempo para temperaturas mínimas, medias y máximas; así como la variable “Muertes” con desplazamiento de 5 semanas.

4. Información Epidemiológica Recopilada

Una página Web del CDC (<http://www.cdc.gov/flu/weekly>) proporciona resúmenes por naciones y regiones con periodicidad semanal para la semana en curso, así como los datos acumulados para una estación dada del año. El sistema de reporte con un seguimiento rápido de las muertes asociadas a la influenza se consigue por medio de dos sistemas: 1. **Sistema de reporte de mortalidad para 122 ciudades en EE.UU.** y 2. **Vigilancia para la influenza-asociada con mortalidad pediátrica** (para personas menores de 18 años de edad). Relativo al primer sistema, cada semana, las oficinas de estadísticas vitales (el equivalente mexicano de nuestras oficinas de registro civil) reportan el número total de certificados de defunción, así como el número de aquellos certificados para los que se declara como causa de la defunción (o la que contribuye al deceso) a la neumonía y a la influenza. Se utiliza un modelo de regresión con periodicidad, el cual se aplica a los datos de los cinco años anteriores. Utilizando un nivel de confianza unilateral del 5% y el modelo de la distribución normal, un aumento de 1.645 desviaciones estándar por encima de la base de muerte por neumonía e influenza se considera, por definición como el *umbral epidémico*. Los datos epidemiológicos utilizados para el presente trabajo se resumen en las variables: “Por ciento de todas las muertes”, “número de semana” y fecha correspondiente. Los datos para las variables climatológicas se presentan en forma sincronizada con los de las variables epidemiológicas, con la finalidad de realizar cálculos de correlaciones (autocorrelaciones y correlaciones cruzadas) para las series de tiempo analizadas. Al respecto, se presentan en forma gráfica los así llamados *correlogramas*.

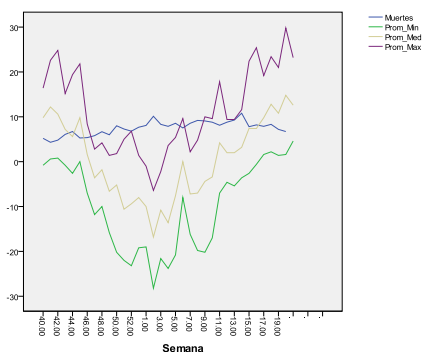


Fig. 8. Series de tiempo para las variables Temperatura Mínima, Temperatura Media, Temperatura Máxima y Muertes

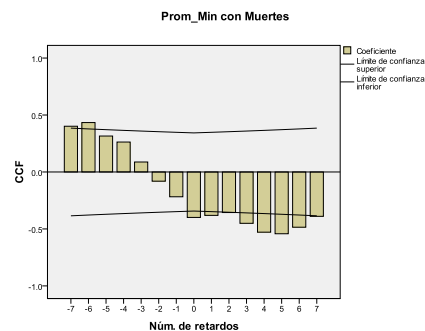


Fig. 9. Autocorrelaciones cruzadas entre muertes y temperaturas mínimas en el estado de Wisconsin. Desplazamientos desde -7 hasta +7 semanas.

De la Fig. 8, no es fácilmente aparente el grado de asociación que la variable “Muertes” pudiera tener con las variables climáticas, ya sea de manera sincronizada o con desplazamientos en el tiempo hacia atrás o hacia delante. Por tales razones, se realizan análisis de autocorrelaciones y de correlaciones cruzadas, de las cuales se presenta una gráfica. Ver Fig. 9. De la misma figura, puede apreciarse que la magnitud máxima de la correlación ocurre para un desplazamiento de +5 semanas y que la correlación es negativa, indicando con esto que, a temperaturas mínimas más bajas, se tendrá una repercusión de un número mayor de muertes 5 semanas después. Para desplazamientos mayores a 5 semanas,

las magnitudes de las correlaciones siguen un patrón decreciente y lo mismo puede apreciarse para desplazamientos menores a 5 semanas.

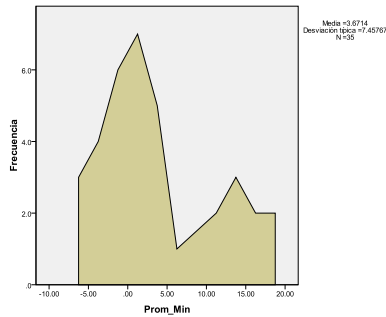


Fig. 10. Distribución de frecuencias de temperaturas mínimas para el estado de Texas, abarcando desde las semanas 40 de 2008 hasta la 20 de 2009.

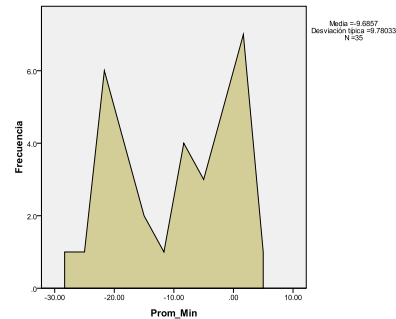


Fig. 11. Distribución de frecuencias de temperaturas mínimas para el estado de Wisconsin, abarcando desde las semanas 40 de 2008 hasta la 20 de 2009.

REFERENCIAS

<http://www.cdc.gov/h1n1flu/> Página Web de CDC (Centers for Disease Control and Prevention)

http://www.who.int/csr/don/2009_06_19/en/index.html Página Web de WHO (World Health Organization)

<http://www.noaa.gov/> Página Web de NOAA (National Oceanic and Atmospheric Administration)

<http://www.ncdc.noaa.gov/oa/ncdc.html> Página Web de NCDC (National Climatic Data Center)

Algunos Modelos Estadísticos Simples para Describir la Evolución de un Proceso Epidemiológico

Ismael Sosa Galindo¹, Luis Cruz-Kuri², Agustín Jaime García Banda³

¹Facultad de Ciencias Administrativas y Sociales. U.V. isoga77@yahoo.com.mx

²Instituto de Ciencias Básicas. U.V. kruz1111@yahoo.com.mx

³Facultad de Ciencias Administrativas y Sociales. U.V. jaimegarciabanda@yahoo.com

1. Introducción

Con el reciente brote de influenza, ahora denominada A(H1N1), gran parte del público en general, así como de los especialistas, empezaron a interesarse en el desarrollo y evolución de este proceso de contagio. En nuestro reporte, se presentan algunos modelos estadísticos simples para tratar de ajustar las series observadas de tiempo correspondientes al número de casos acumulados por contagio, así como al número de defunciones acumuladas, como consecuencia de haber contraído esta enfermedad. Los modelos aquí propuestos son muy conocidos, tales como el de regresión lineal simple, el de regresión exponencial y el logístico, entre otros. Debe ser obvio que el primero de estos modelos es el más sencillo pero su validez de ajuste es a muy corto plazo; en tanto que los siguientes dos modelos tienen una vigencia a un plazo mayor. Las series reales de tiempo de la evolución de la presente influenza, para los correspondientes números acumulados de casos de contagio y de defunciones, se han recabado por nosotros, con información obtenida directamente del CDC (Centers for Disease Control and Prevention) de Atlanta, para el caso de EE.UU.; lo anterior se realizó a nivel de todo el país y también con desglose a nivel estatal. Para el caso de México, así como de otros países, los datos relevantes fueron obtenidos a través de la WHO (World Health Organization, por sus siglas en inglés).

2. Obtención de la Información y Análisis

Para los datos de la nueva influenza A(H1N1) se recurrió a las fuentes primarias de el CDC, así como a la Organización Mundial de la Salud (WHO por sus siglas en inglés). En particular, la agencia WHO prepara reportes con actualizaciones frecuentes, diarias y en algunos casos dos veces al día, del estado de esta epidemia en relación al número de casos acumulados en una fecha dada, así como el número de defunciones acumuladas hasta ese momento y por otra parte los nuevos casos y las nuevas defunciones. El tipo de información primaria que se utilizó para el presente trabajo es como el que aparece en la Tabla 1, que se incluye enseguida. Es pertinente mencionar que, por consideraciones de espacio, solo se incluyen algunos países y territorios, pero los datos en el apartado correspondiente a “Gran Total” sí incluyen a todos los países, que a la fecha del reporte tienen al menos un caso. Igualmente la Tabla 1, solo representa la situación en un momento fijo en el tiempo; de tal manera que si se quisiera tener un seguimiento completo, desde que empezaron a registrarse casos de influenza, se requerirían tantas tablas como actualizaciones se abarquen, lo cual al momento de cerrar nuestro reporte da un total de 51 actualizaciones.

Tabla 1*. Casos Confirmados por Laboratorio de la Nueva Influenza A(H1N1) Reportados Oficialmente a la Asociación WHO por Países Territorios y Areas Según el Reglamento Internacional de Salud (2005). Actualización 51 de Fecha 19 de Junio de 2009

| País, Territorio y Area | Total Acumulado | | Nuevos Casos Confirmados desde el Ultimo Periodo de Reporte | |
|-------------------------|-----------------|-------------|---|-------------|
| | Casos | Defunciones | Casos | Defunciones |
| Argentina | 918 | 4 | 185 | 3 |
| Australia | 2199 | 0 | 87 | 0 |
| Canadá | 4905 | 12 | 856 | 5 |
| Chile | 3125 | 2 | 790 | 0 |
| China | 519 | 0 | 137 | 0 |
| España | 512 | 0 | 13 | 0 |
| Estados Unidos | 17855 | 44 | 0 | 0 |
| Japón | 690 | 0 | 24 | 0 |
| México | 7624 | 113 | 1383 | 5 |
| Reino Unido | 1752 | 1 | 291 | 0 |
| Tailandia | 518 | 0 | 208 | 0 |
| Gran Total ** | 44287 | 180 | 4667 | 13 |

FUENTE: Organización Mundial de la Salud (WHO por sus siglas en inglés)

*Selección de algunos países y territorios

**El Gran Total corresponde a todos los países afectados a la fecha

3. Procesamiento de la Información

Se crearon para cada país los archivos correspondientes, los cuales fueron analizados con el programa de cómputo estadístico SPSS 17. Algunas series de tiempo se estudiaron en forma individual, en tanto que en otros casos se incluyeron dos o más series de tiempo simultáneamente. Por consideraciones de espacio, en el presente trabajo no se incluyen los modelos de regresión, aunque se hace mención que el más satisfactorio fue el de regresión logística.

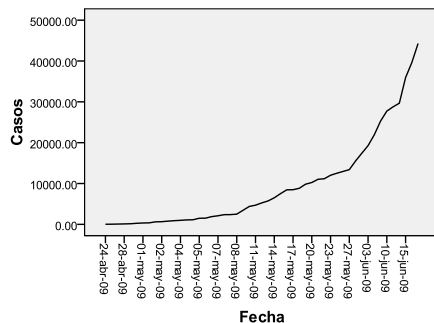


Fig. 1. Serie de Tiempo de casos acumulados de influenza A(H1N1). Global. Todos los Países

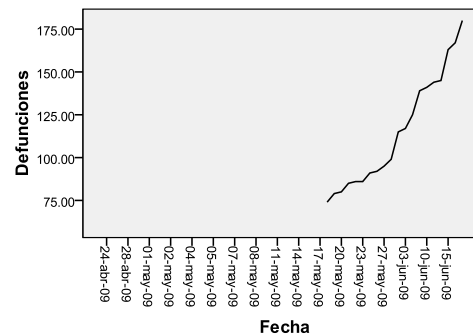


Fig. 2. Serie de Tiempo de defunciones acumuladas por influenza A(H1N1). Global. Todos los Países

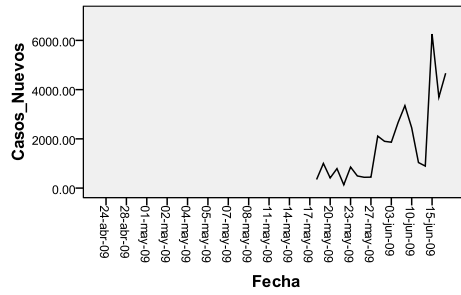


Fig. 3. Serie de Tiempo de casos nuevos de influenza A(H1N1). Global. Todos los Países.

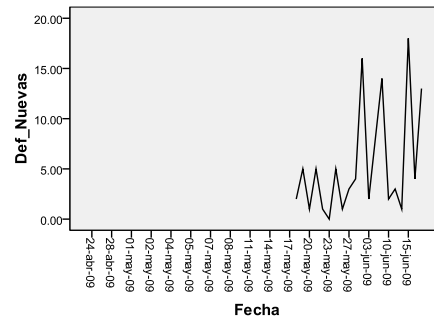


Fig. 4. Serie de Tiempo de defunciones nuevas en el periodo correspondiente atribuidos a influenza A(H1N1). Global. Todos los Países.

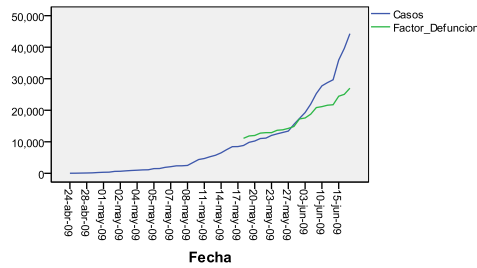


Fig. 5. Series de tiempo para casos acumulados y defunciones acumuladas (factor de escala:150) por influenza A(H1N1). Global. Todos los países.

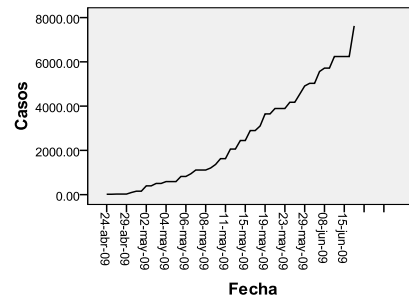


Fig. 6. Serie de Tiempo de casos acumulados de influenza A(H1N1). México.

De la gráfica (Fig. 5) puede apreciarse que la tasa de mortalidad es mayor para fechas anteriores al 28 de mayo de 2009 que para las fechas posteriores, en que la tasa se ha reducido.

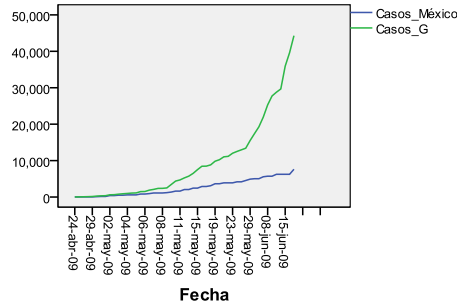


Fig. 7. Series de tiempo de casos acumulados para México y para todos los países.

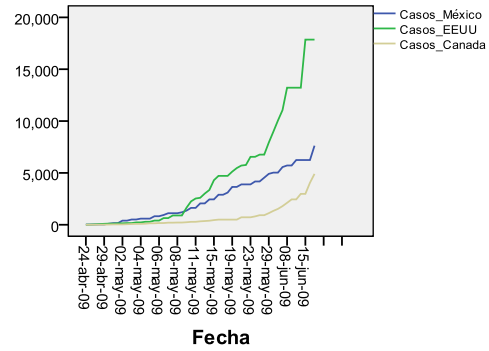


Fig. 8. Evolución de la epidemia en tres países. Casos acumulados para México, Canadá y EE.UU.

La variable “Casos_México” se refiere al país que se indica y la variable “Casos_G” se refiere (Fig.7) a todos los países. La gráfica en la Fig. 8 describe la evolución de la epidemia en forma simultánea para tres países. Nótese que aunque EE.UU. inició por debajo de México, más adelante fue rebasado por el primero. Canadá se ha mantenido por debajo de México.

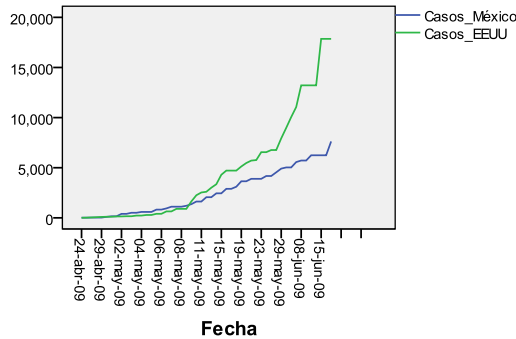


Fig. 9. Evolución de la epidemia en dos países. Casos acumulados para México y para EE.UU.

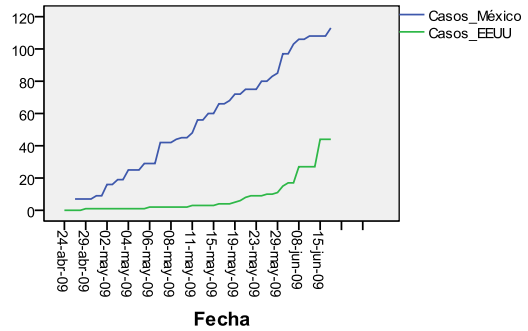


Fig. 10. Evolución de la epidemia en dos países. Defunciones acumuladas para México y para EE.UU.

Aunque EE.UU. inició con un número menor de casos en las fechas tempranas, más adelante rebasó al número de casos de México por varias razones que requieren estudiarse con mayor detalle. Una de ellas es mayor tamaño de población para EE.UU, otra puede ser: distinto manejo epidemiológico. Por otra parte, aunque (ver Fig.10) el número de casos en EE.UU. supera (ver Fig.9) en mucho al correspondiente para México, los papeles se invierten en lo correspondiente a defunciones por una variedad de causas que requieren estudiarse; una de ellas puede atribuirse a la diferencia en estilo de vida, otra a la diferencia en atención médica una vez que se contrae la enfermedad.

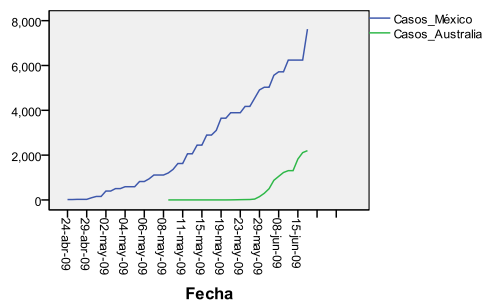


Fig. 11. México vs Australia. Casos acumulados.

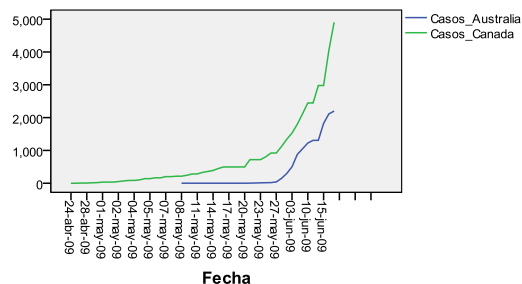


Fig. 12. Australia vs Canadá. Casos acumulados.

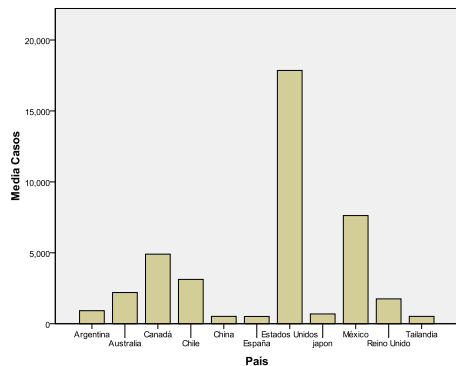


Fig. 13. Países con al menos 500 casos de influenza A(H1N1) acumulados al 19 de junio de 2009 (actualización 51).

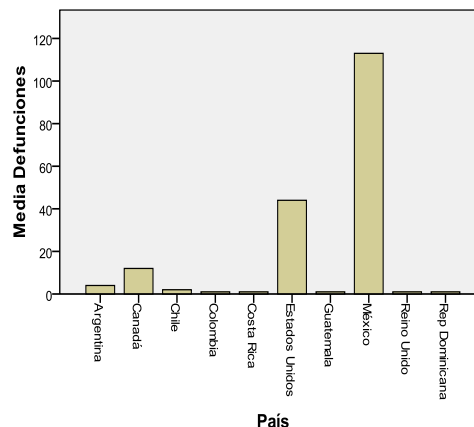


Fig. 14. Países con al menos una defunción acumulada a la fecha 19 de Junio de 2009 (actualización 51).

Puede apreciarse de la gráfica (Fig.13) que el país que más casos tiene hasta esa fecha es Estados Unidos y que le siguen México, Canadá, Chile, Australia, Reino Unido, para citar a los seis con mayor número de casos. De la gráfica (Fig. 14) puede observarse que el país con el mayor número de defunciones acumuladas a la fecha es México, y le siguen Estados Unidos, Canadá, Argentina y Chile, en ese orden.

REFERENCIAS

- http://www.who.int/csr/don/2009_06_19/en/index.html [para varias actualizaciones, desde la número 51 al 19 de Junio de 2009]
- <http://www.cdc.gov/h1n1flu/espanol/>

Estudio Morfométrico de la Tortuga *Kinosternon Herrerae* en un Arrollo Situado en el Municipio de Xalapa, Ver.

Gustavo Aguirre León¹, Erasmo Cázares Hernández², Blanca Hervier Martínez³

Universidad Veracruzana

¹gustavo.aguirre@inecol.edu.mx

²caeh72@hotmail.com

³bhervier@hotmail.com

Resumen

Introducción:

México cuenta con la mayor diversidad de reptiles del planeta y en proporción al territorio, la fauna de tortugas de México puede ser considerada como una de las más diversas en el mundo (Smith y Smith, 1979). En el estado de Veracruz se distribuyen 12 especies de tortugas de agua dulce de las familias Kinosternidae, Staurotypidae, Dermatemyidae, Chelydridae, Emydidae y Bataguridae.

Kinosternon herrerae es una especie de tortuga dulceacuícola endémica del este de México y se encuentra ubicada en la categoría de "protección especial" en la legislación ambiental mexicana (NOM-059-ECOL-2001). La última lista roja de especies amenazadas de la IUCN la consideran casi amenazada. En el municipio de Xalapa, Veracruz, *K. herrerae* es el único representante del orden Testudines. La transformación de los ambientes que ocupa *K. herrerae* ha provocado una disminución en sus poblaciones (Ramírez-Bautista *et al.*, 1993), lo cual hace necesario conocer más sobre su biología y ecología para poder identificar medidas de conservación de la especie.

Objetivo:

Analizar las características morfométricas y de dimorfismo sexual, así como la proporción de sexos de una población de *Kinosternon Herrerae*.

Metodología:

La población de *Kinosternon herrerae* estudiada pertenece al arroyo "La Bomba", (96°52'49" W, 19°31'25" N). Tiene una anchura máxima de unos 250 cm, y sus zonas más angostas son de unos 150 cm. No sobrepasan los 100 cm de profundidad.

Para la captura de ejemplares se llevaron a cabo 25 sesiones de trampeo, dos o tres por mes, de Octubre de 1996 a Octubre de 1997. Se utilizaron cuatro trampas construidas con malla gallinera descritas por Iverson (1979), cebadas con latas de sardina. Se marcó a cada nuevo individuo capturado para su identificación en recapturas posteriores.

Para determinar si la proporción sexual es diferente de 1:1 se utilizó la prueba de Chi-cuadrado para dos categorías. Para este análisis se tomaron en cuenta sólo individuos con caracteres sexuales secundarios evidentes o incipientes.

Se registraron 49 variables morfométricas incluido el peso (Figura 1). Las diferencias entre algunas de las medidas más importantes de machos y hembras se analizaron con una prueba de t de dos colas.

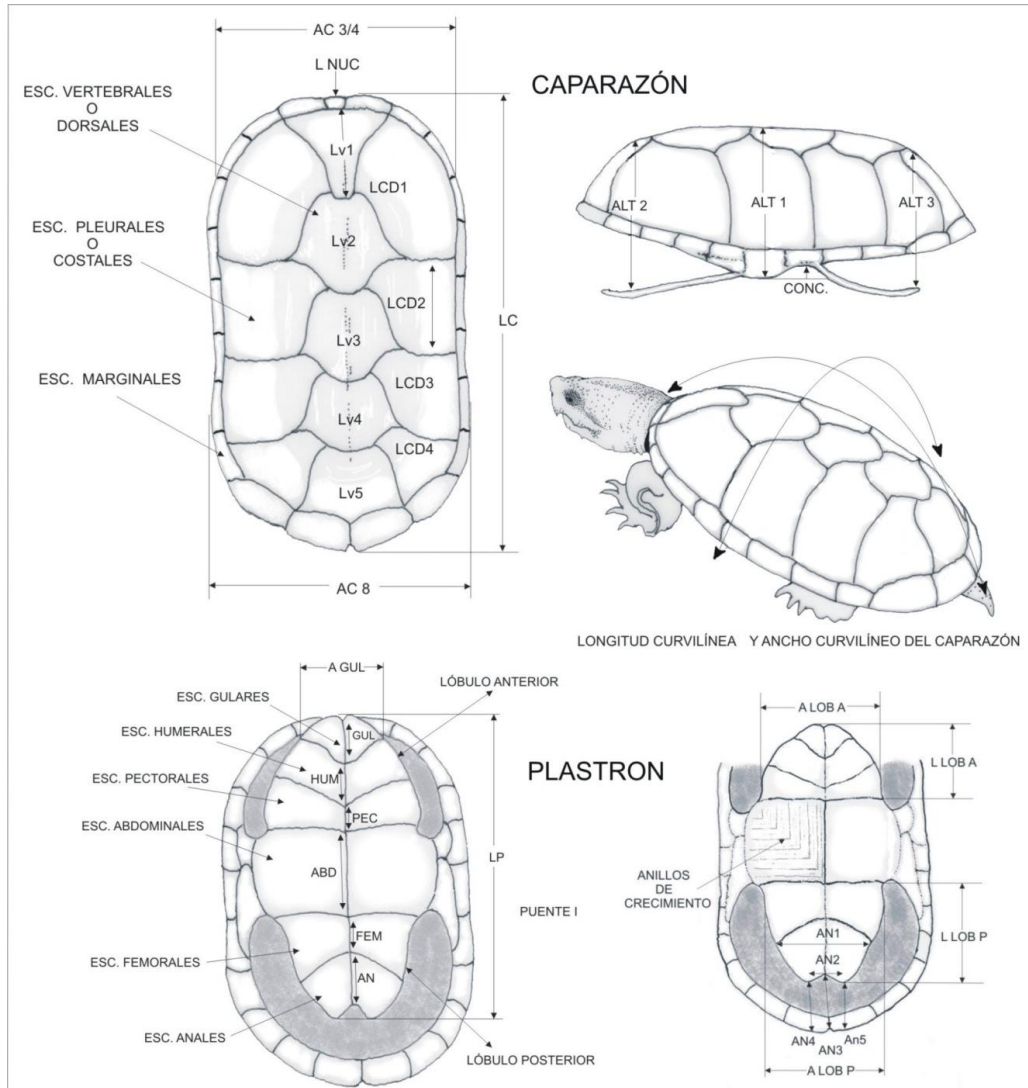


Figura 1. Variables morfométricas registradas durante el estudio (los anillos de crecimiento se miden siguiendo el mismo sentido longitudinal en el cual se mide ABD).

El dimorfismo sexual se evaluó con el índice de dimorfismo sexual (IDS) de Lovich y Gibbons (1992), dado por el cociente del tamaño promedio del sexo más grande entre el del sexo de menor talla. El valor del IDS se considera positivo cuando las hembras son más grandes que los machos, y negativo en la situación contraria.

Resultados

De una muestra de 80 individuos, se confirmó el sexo de 27 machos y 46 hembras; esta muestra contuvo tres individuos considerados como machos y dos como hembras menores de 110 mm (LC) a juzgar por el desarrollo incipiente de caracteres sexuales secundarios. Las hembras son significativamente más numerosas que los machos ($X^2 = 4.438$, $P > 0.05$, g.l. = 1). La proporción es de 1.7 hembras por macho.

Los machos son significativamente más grandes que las hembras en la mayoría de las dimensiones corporales (Tabla 1). La talla máxima (LC) de machos fue de 165 mm, y 157 mm en hembras, y el peso máximo de machos fue de 780 g y el de hembras 580 g. Se puede encontrar una descripción estadística de los datos más detallada en el resumen de la contribución oral en la que se expuso un trabajo sobre simplificación de los estudios morfométricos de *Kinosternon herrerai* en este mismo encuentro.

Tabla 1. Diferencias morfométricas de *Kinosternon herrerai*, medidas (mm) de machos menos medidas de hembras. (x = media); t= estadístico de t-Student; gl= grados de libertad; p= probabilidad de error considerándose significativo con $p < 0.05$.

| Caparazón | x | t | gl | p | Plastro | x | t | gl | p |
|-----------|------|--------|----|---------|---------|------|-------|----|---------|
| LC | 16.4 | -7.05 | 44 | <0.0001 | LP | 6.9 | -3.69 | 55 | <0.05 |
| ALT 1 | 5.8 | -4.03 | 31 | <0.05 | A LOB | 4.8 | -4.46 | 31 | <0.0001 |
| ALT 2 | 4.1 | -4.03 | 44 | <0.05 | A LOB | 2.8 | -3.01 | 42 | <0.05 |
| ALT 3 | 3.0 | -2.74 | 43 | <0.05 | GUL | 0.36 | -0.66 | 41 | >0.05 |
| LC curv | 19.1 | -6.35 | 34 | <0.0001 | HUM | 0.9 | -1.71 | 39 | >0.05 |
| AC curv | 10.3 | -3.93 | 32 | <0.05 | PEC | 1.9 | -2.66 | 58 | <0.05 |
| LONG | 0.9 | -2.38 | 27 | <0.05 | ABD | 2.0 | -2.36 | 37 | <0.05 |
| CAB A | 4.7 | -4.0 | 28 | <0.05 | FEM | 0.9 | -1.77 | 61 | >0.05 |
| CAB L | 9.4 | -6.10 | 36 | <0.0001 | AN | 2.5 | -3.59 | 39 | <0.05 |
| PUENTE D | 0.33 | -0.54 | 41 | >0.05 | CONC | 2.2 | -8.30 | 25 | <0.0001 |
| PUENTE I | 0.27 | -0.420 | 41 | >0.05 | PESO | 92.5 | -3.67 | 31 | <0.05 |

El perfil dorsal del caparazón en las hembras es oval y ligeramente expandido, y el de los machos presenta una pendiente pronunciada (Figura 2).

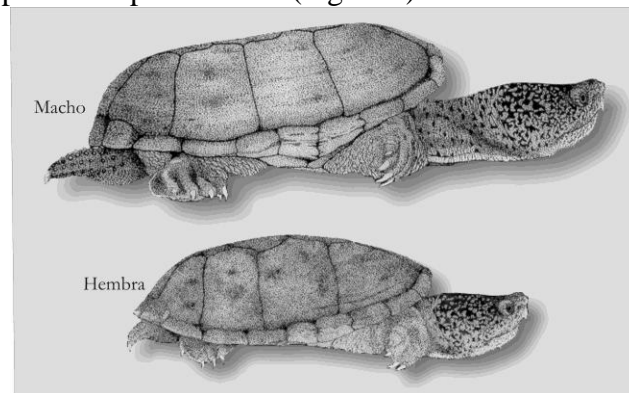


Figura 2. Detalle del perfil de machos y hembras de *Kinosternon herrerai*. Arroyo “La Bomba”, Mpio. de Xalapa, Ver. (Dibujo: Erasmo Cázares Hernández).

Conclusiones:

En este estudio se obtuvo una proporción de 1.7 ♀♀: 1 ♂♂. En contraste, Carr y Mast (1988) obtuvieron para esta especie una proporción no significativamente diferente de 1:1 con una muestra más pequeña (7 machos y 10 hembras). Existen varios reportes de proporciones desiguales de sexos en poblaciones de tortugas de agua dulce. A veces la proporción en una población puede variar estacionalmente (Morreale *et al.*, 1984; Dodd *et al.*, 1988; Mitchell, 1988) o incluso año con año en función de la técnica de muestreo utilizada. A pesar de esto la proporción puede estar efectivamente sesgada debido a

migraciones (Auffenberg y Iverson, 1979), diferencias de comportamiento entre sexos, la edad o talla en que se alcanza la madurez sexual o una combinación de estos factores (Gibbons y Lovich, 1990). También podrían estar influyendo las condiciones ambientales durante la incubación, ya que es una especie con determinación del sexo por temperatura (Iverson, 1998).

Kinosternon herrerae es una de las especies con dimorfismo sexual marcado (Carr y Mast, 1988). Estos autores obtuvieron un cociente ♂/♀ de -1.15 en LC, y de -1.06 en LP, similar a los obtenidos en nuestro estudio (LC = -1.13 y LP = -1.06). Sin embargo el tamaño máximo (LC) de los machos y hembras reportados aquí (165 mm y 157 mm respectivamente) difiere de otros estudios. Smith y Smith (1979) reportan un máximo de 170 mm en machos y 150 mm en hembras y Carr y Mast (1988) 172 mm y 148 mm.

La expresión de dimorfismo sexual acentuado en tortugas es muy común, ya sea a favor de un sexo u otro. En el caso de la familia Kinosternidae aproximadamente el 80 % de las especies los machos son más grandes que las hembras, con algunas variaciones entre poblaciones (Berry y Shine, 1980; Gibbons y Lovich 1990; Iverson, 1985). Hay diversos factores biológicos que pueden influir en el dimorfismo sexual entre poblaciones de una especie. Ciertas teorías sugieren que es el resultado de fuerzas de selección natural, debido a la interacción diferencial de cada sexo con su ambiente. Se predice mayor dimorfismo en poblaciones limitadas por alimento (teoría de partición del nicho: Schoener, 1967; 1968), y en poblaciones con pocos congéneres y competidores (teoría de la competencia: Schoener, 1977). En el caso de la población analizada el dimorfismo puede estar asociado con las características del hábitat: al tratarse de un arroyo de proporciones reducidas, los machos tienen mayores probabilidades de defender un territorio que en un ambiente abierto como un humedal.

Este trabajo da una idea de la gran variabilidad que existe entre poblaciones, influenciada por factores ambientales y genéticos. Es necesario realizar más estudios que permitan comprender la interacción e importancia de tales factores.

REFERENCIAS

- Auffenberg, W., Iverson, J. B. (1979). Demography of terrestrial turtles. *In*: M. Harless, y H. Morlock (Eds.). *Turtles. Perspectives and Research*. Wiley Interscience, New York. Pp. 541-569.
- Berry, J. F., Shine, R. (1980). Sexual size dimorphism and sexual selection in turtles (Order Chelonia). *Oecologia* 44, 185-191.
- Carr, J. L., Mast, R. B. (1988). Natural history observations of *Kinosternon herrerae* (Testudines: Kinosternidae). *Trianea* 1, 87-97.
- Dood, C. K. Jr., Enge, K. M., Stuart, J. N. (1988). Aspects of the biology of the flattened musk turtle, *Sternotherus depressus* in northern Alabama. *Bulletin of the Florida State Museum, Biological Sciences Series* 34, 1-64.
- Gibbons, W. J., Lovich, J. E. (1990). Sexual dimorphism in turtles with emphasis on the slider turtle (*Trachemys scripta*). *Herpetological Monographs* 4, 1-29.
- Iverson, J. B. (1979). Another inexpensive turtle trap. *Herpetological Review* 10, 55.
- Iverson, J. B. (1985). Geographic variation in sexual dimorphism in the mud turtle *Kinosternon hirtipes*. *Copeia* 1985, 388-393.
- Iverson, J. B. (1998). Molecules, Morphology, and Mud Turtle Phylogenetics (Family Kinosternidae). *Chelonian Conservation and Biology* 3, 113-117.

- Lovich, J. E., Gibbons, J. W. (1992). A review of techniques quantifying sexual size dimorphism. *Growth, Development and Aging* 56, 269-281.
- Mitchell, J. C. (1988). Population ecology and life histories of the freshwater turtles *Chrysemys picta* and *Sternotherus odoratus* in an urban lake. *Herpetological Monographs* 2, 40-61.
- Morreale, S. J., Gibbons, J. W., Congdon, J. C. (1984). Significance of activity and movement in the yellow-bellied slider turtle (*Pseudemys scripta*). *Canadian Journal of Zoology* 62, 1038-1042.
- Ramírez-Bautista, A., González-Romero, A., López-González, C. A. (1993). Estudio preliminar de la Herpetofauna del Municipio de Xalapa. In: R. López-Moreno (Ed.). Ecología urbana aplicada a la ciudad de Xalapa. Instituto de Ecología. Xalapa, Veracruz. Pp. 165-188.
- Ream, C., Ream, R. (1966). The influence of sampling method on the estimation of population structure in painted turtles. *The American Midland Naturalist* 75, 325-338.
- Schoener, T. W. (1967). The ecological significance of sexual dimorphism in size in the lizard *Anolis conspersus*. *Science* 155, 474-477.
- Schoener, T. W. (1968). The *Anolis* lizards of Bimini: Resource partitioning in a complex fauna. *Ecology* 49, 704-726.
- Schoener, T. W. (1977). Competition and the niche. v.7A. In: Gans, C. y Tinkle, D. W. (Eds.). *Biology of the Reptilia*. Academic, London. Pp. 35-136
- Smith, H. M., Smith, R. B. (1979). Synopsis of the Herpetofauna of México, Guide to mexican turtles bibliographic addendum III. Volume VI. Pp. 151-154.

Evaluación Sensorial de Aderezos Ricos en Ácidos Grasos Omega-3 y Omega-6

Martínez-Alejo J.M.¹, García-Barradas O.², Mendoza-López M. R.³, Muñoz-Muñiz O.D.,⁴ Jiménez-Fernández M.⁵

¹Instituto de Ciencias Básicas. Universidad Veracruzana. juanm_alejo@yahoo.com.mx

²Unidad de Servicios de Apoyo en Resolución Analítica. Universidad Veracruzana
osgarcia@uv.mx

³Unidad de Servicios de Apoyo en Resolución Analítica. Universidad Veracruzana

⁴Unidad de Servicios de Apoyo en Resolución Analítica. Universidad Veracruzana
omuniz@uv.mx

⁵Instituto de Ciencias Básicas. Universidad Veracruzana. marijife@hotmail.com

Resumen

Actualmente, las enfermedades cardiovasculares son la principal causa de muerte a nivel mundial. Se ha comprobado que el incorporar en la dieta ácidos grasos poliinsaturados omega-3 y omega-6 puede ayudar a prevenir este padecimiento. El objetivo de este trabajo fue realizar dos aderezos variando la proporción de ácidos grasos omega-3 y omega-6, uno con una mayor proporción de ácidos grasos omega-3 y otro con una mayor proporción de ácidos grasos omega-6; y evaluar su calidad sensorial. Para la evaluación sensorial se utilizó una escala hedónica de 9 puntos. El análisis de varianza (ANOVA) de los resultados permitió comprobar que los panelistas fueron capaces de diferenciar los dos aderezos, mostrando un mayor agrado hacia el aderezo rico en ácidos grasos omega-6 (5.5 ± 1.7) que hacia el aderezo rico en ácidos grasos omega-3 (3.5 ± 1.9). Se observó que esta preferencia no se vio influenciada por el sexo de los panelistas, ya que tanto los panelistas del sexo femenino ($n=39$, $p < 0.05$) como los del sexo masculino ($n=24$, $p < 0.05$) coincidieron en su preferencia. Por otro lado, se pudo observar también que los panelistas que acostumbran consumir aderezos de manera frecuente y poco frecuente pudieron distinguir fácilmente entre nuestros productos ($p < 0.05$), sin embargo, las personas que casi nunca consumen aderezos no encontraron diferencias entre ellos ($p > 0.05$).

Palabras clave: Aderezo, Ácidos grasos, Omega-3, Omega-6, evaluación sensorial.

Introducción

Según estimaciones hechas por la Organización Mundial de la Salud (OMS), la principal causa de muertes a nivel mundial en el año de 2005 fueron las enfermedades cardiovasculares (WHO, 2008). Sin embargo, la evidencia científica de las últimas décadas avala el papel positivo que una alimentación saludable tiene en la prevención de las enfermedades coronarias (Carrero *et al.*, 2006).

Aunque los lípidos son culpables en cierta forma del aumento de las enfermedades cardiovasculares, diversos estudios han comprobado que el riesgo de padecer enfermedades cardiovasculares está claramente relacionada con el tipo de grasa que se ingiera en la dieta, más que en la cantidad de la misma. Se ha comprobado que los ácidos grasos poliinsaturados omega-3 (ácido α -linolénico, ácido eicosapentaenoico y ácido

docosahexaenoico) reducen los niveles plasmáticos de triglicéridos, las arritmias cardiacas y el riesgo de padecer enfermedades coronarias, mientras que los ácidos grasos omega-6 (principalmente el ácido linoleico) ayudan a reducir la concentración plasmática de las lipoproteínas de baja densidad (LDL) ó el comúnmente denominado colesterol malo, además de aumentar el metabolismo de los lípidos. Sin embargo, diversos estudios han denotado que la proporción de ácidos grasos omega-6/omega-3 que nosotros incorporamos en la dieta es un factor importante para tener una buena salud cardiovascular (Simopoulos, 2004).

Debido a esto, la misma OMS recomienda sustituir la grasa saturada por monoinsaturada, aumentar la ingesta de frutas y verduras para conseguir niveles óptimos de vitaminas y antioxidantes así como incluir en la dieta ácidos grasos poliinsaturados omega-3 y omega-6 (WHO, 2008).

Por todo lo anterior, es necesario diseñar y elaborar alimentos que además de nutrir disminuyan el riesgo de padecer estas enfermedades. El objetivo de este estudio fue realizar dos aderezos variando la proporción de ácidos grasos omega-3 y omega-6, uno con una proporción 2:1 de ácidos grasos omega-3 contra omega-6, y otro con la proporción recomendada 2:1 de ácidos grasos omega-6 con respecto a omega-3; y evaluar su calidad sensorial.

Material y Métodos

Materia prima.

Las especias utilizadas para la elaboración de los aderezos fue adquirida bajo las marcas de Catarinos y Escosa. El aceite utilizado para la elaboración del aderezo con una proporción 2:1 de omega-6/omega-3, fue aceite de canola refinado libre de antioxidantes, proporcionado por la empresa Industrial Patrona S.A. de C.V. El aceite utilizado para la elaboración del aderezo con una proporción 1:2 de omega-6/omega-3, fue aceite de linaza extravirgen libre de antioxidantes, adquirido por la empresa Olimu.

Elaboración de los aderezos.

Para realizar los aderezos, se realizaron pruebas piloto variando el tipo, número y cantidad de especias incorporadas en la formulación para identificar aquella que presentaba mejores características sensoriales. La formulación final del aderezo incluyó laurel, albahaca, romero, orégano, estragón, cebollín, pimienta, ajo.

Evaluación sensorial.

Para la evaluación sensorial se contó con un panel de 63 personas no entrenadas y se utilizó una escala hedónica de 9 puntos (1= me disgusta extremadamente, 2= me disgusta mucho, 3= Me disgusta moderadamente, 4= me disgusta ligeramente, 5= Ni me gusta ni me disgusta, 6= Me gusta ligeramente, 7= Me gusta moderadamente, 8= Me gusta mucho, 9= me gusta extremadamente).

Análisis estadístico.

Con los resultados de las evaluaciones se construyeron bases de datos en el paquete estadístico Statistica 2004 (StatSoft, Inc. (2004). STATISTICA (data analysis software

system), versión 7. www.statsoft.com.) calculándose promedios y desviaciones estándar, además de realizarse un ANOVA.

Resultados

Análisis preliminar

En la figura 1 podemos observar las características con las que contó el panel evaluador. En cuanto al género de los panelistas que realizaron la evaluación, el 38% de ellos fueron hombres mientras que el 62% fueron mujeres. Un 28% de ellos aseguraron tener una edad entre los 11 y 20 años, un 70% mencionaron encontrarse en el rango de edad de 21 a 30 años, y solamente un 2% de ellos tuvieron una edad de entre 31 a 40 años. En cuanto a la frecuencia de consumo, 3% de ellos mencionó que lo consumen muy frecuentemente (una vez cada tres días), el 33% mencionó que lo acostumbra consumir de manera frecuente (una vez a la semana), otro 33% mencionó que consume aderezos de manera poco frecuente, es decir, una vez cada mes, mientras que el 24% de los panelistas aseguró que casi nunca consume aderezos.

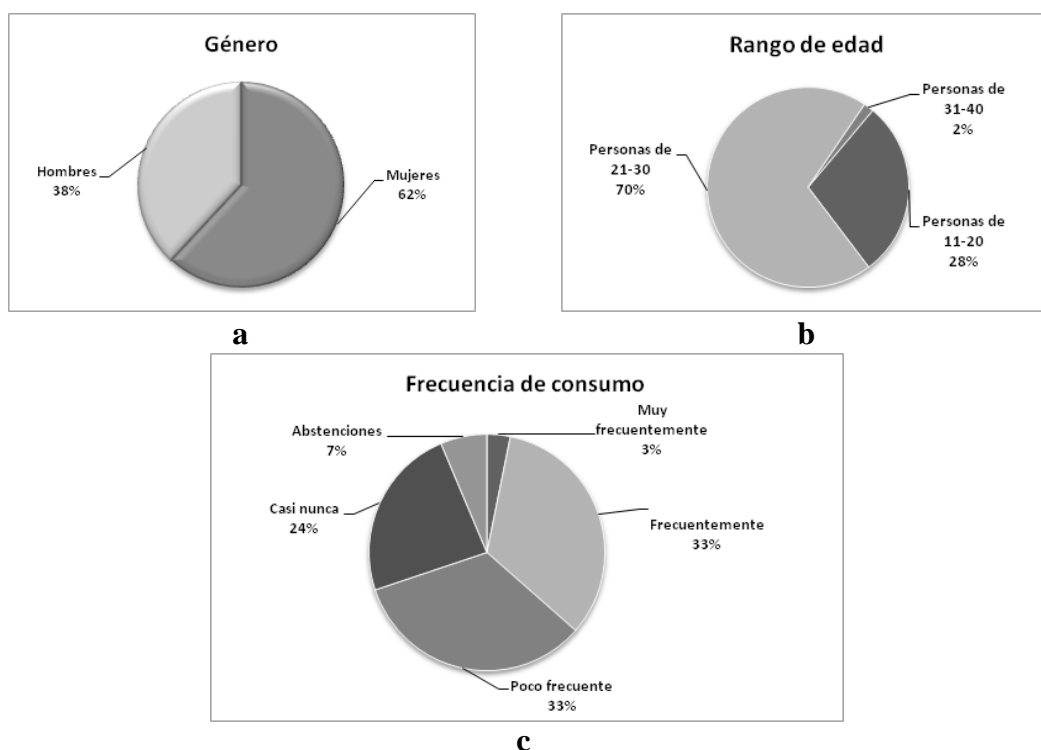


Figura 1. a) Género de los panelistas; b) Rango de edad de los panelistas; c) Frecuencia de consumo de aderezos de los panelistas

Análisis estadístico de los aderezos

El análisis de los resultados se llevó a cabo con el programa Statistica 7, mediante un ANOVA de una sola vía planteándose las siguientes hipótesis:

H₀ = la μ de los aderezos son iguales; **H_a** = los aderezos son distintos

El análisis estadístico lo presentamos de la siguiente manera:

- **Análisis general:** en este análisis no se tomó en cuenta el género, la edad y la frecuencia de consumo de los panelistas.

- **Análisis por género:** para llevar a cabo este análisis se separaron los 63 panelistas por sexo. Del total de panelistas, 39 fueron mujeres y 24 hombres.
- **Análisis por frecuencia de consumo:** para este análisis, se juntó en un solo grupo a los panelistas que consumen de manera muy frecuente y frecuentemente aderezos (23 panelistas), en otro grupo los que consumen aderezos de manera poco frecuente (21 panelistas), y en un último grupo los panelistas que casi nunca consumen aderezos (15 personas).

- **Análisis General.**

El resultado del ANOVA para todo nuestro panel evaluador dio un valor de p menor al valor de $\alpha(0.05)$, lo que indica que los panelistas en general encontraron diferencias significativas en cuanto al sabor de los dos aderezos (Tabla 1). La calificación para el aderezo rico en omega-6 fue de 5.51 ± 0.23 , mientras que para el aderezo rico en omega-3 fue de 3.5 ± 0.23 .

Tabla 1. ANOVA de la calificación de los panelistas en general.

| | SS | Grados de libertad | MS | F-ratio | P-value |
|-------------------|----------|--------------------|----------|----------|----------|
| Intercepto | 2553.602 | 1 | 2553.602 | 728.4192 | 0.000000 |
| "Var1" | 124.034 | 1 | 124.034 | 35.3809 | 0.000000 |
| Error | 431.198 | 123 | 3.506 | | |

- **Análisis por género.**

Femenino.

El análisis de las panelistas incluidas en el estudio nos mostró que éstas fácilmente pueden distinguir sensorialmente a los dos aderezos, debido a que el valor de p fue menos que el valor de α de 0.05 (Tabla 2).

Tabla 2. ANOVA de la calificación de los panelistas del sexo femenino.

| | SS | Grados de libertad | MS | F | p |
|-------------------|----------|--------------------|----------|----------|----------|
| Intercepto | 1474.830 | 1 | 1474.830 | 422.0982 | 0.000000 |
| Muestra | 79.765 | 1 | 79.765 | 22.8288 | 0.000009 |
| Error | 262.053 | 75 | 3.494 | | |

Masculino

El análisis de varianza de los resultados obtenidos por los panelistas del sexo masculino nos dio como resultado un valor de p igual a 0.001025, el cual se encuentra también por debajo del valor de α (0.05), lo que indica que para los hombres los aderezos son distintos (Tabla 3).

Tabla 3. ANOVA de la calificación de los panelistas del sexo masculino.

| | SS | Grados de libertad | MS | F | p |
|-------------------|----------|--------------------|----------|----------|----------|
| Intercepto | 1083.000 | 1 | 1083.000 | 302.0798 | 0.000000 |
| Muestra | 44.083 | 1 | 44.083 | 12.2961 | 0.001025 |
| Error | 164.917 | 46 | 3.585 | | |

-

- **Análisis por frecuencia de consumo.**

Frecuentemente.

El valor estadístico de p resultado del ANOVA de los panelistas que aseguraron que consumen aderezos de manera frecuente fue de 0.000032 (Tabla 4). Este valor se encuentra por debajo del valor establecido de α de 0.05, lo que indica que los aderezos son diferentes.

Tabla 4. ANOVA de la calificación de los panelistas que consumen aderezos frecuentemente.

| | SS | Grados de libertad | MS | F | p |
|-------------------|----------|--------------------|----------|----------|----------|
| Intercepto | 809.7609 | 1 | 809.7609 | 246.3854 | 0.000000 |
| Muestra | 70.6304 | 1 | 70.6304 | 21.4907 | 0.000032 |
| Error | 144.6087 | 44 | 3.2866 | | |

Poco frecuentemente.

El valor estadístico de p para el caso de los panelistas que aseguraron que consumen aderezos de manera poco frecuente fue de 0.000536 (tabla 5), el cual, debido a que se encuentra por debajo al valor establecido de α de 0.05, indica que los aderezos son distintos.

Tabla 5. ANOVA de la calificación de los panelistas que consumen aderezos poco frecuentemente.

| | SS | Grados de libertad | MS | F | p |
|-------------------|----------|--------------------|----------|----------|----------|
| Intercepto | 1040.024 | 1 | 1040.024 | 368.3052 | 0.000000 |
| Muestra | 40.024 | 1 | 40.024 | 14.1737 | 0.000536 |
| Error | 112.952 | 40 | 2.824 | | |

Casi nunca.

El ANOVA de los panelistas que aseguraron que casi nunca consumen aderezos nos mostró como resultado un valor estadístico de p de 0.181604 (ver tabla 6), el cual es mayor al valor de α (0.05), lo que indica que para estos panelistas los aderezos fueron iguales.

Tabla 6. ANOVA de la calificación de los panelistas que casi nunca consumen aderezos.

| | SS | Grados de libertad | MS | F | p |
|---------------------|----------|--------------------|----------|----------|----------|
| Intercepto | 625.6333 | 1 | 625.6333 | 121.8766 | 0.000000 |
| Calificación | 9.6333 | 1 | 9.6333 | 1.8766 | 0.181604 |
| Error | 143.7333 | 28 | 5.1333 | | |

Conclusión

El análisis de varianza (ANOVA) de los resultados permitió comprobar que los panelistas son capaces de diferenciar los dos aderezos, mostrando un mayor agrado hacia el aderezo rico en ácidos grasos omega-6 (5.51 ± 0.23) que hacia el aderezo rico en ácidos grasos omega-3 (3.52 ± 0.23). Se observó que esta preferencia no se vio influenciada por el sexo de los panelistas, ya que tanto los panelistas del sexo femenino ($n=39$, $p < 0.05$) como los del sexo masculino ($n=24$, $p < 0.05$) coincidieron en su preferencia. Por otro lado, se pudo observar también que los panelistas que acostumbran consumir aderezos de manera frecuente y poco frecuente pudieron distinguir fácilmente entre nuestros productos ($p < 0.05$); sin embargo, las personas que casi nunca consumen aderezos no encontraron diferencias entre ellos ($p > 0.05$). Los resultados obtenidos sugieren que el aderezo con una proporción mayor de omega-6 es más agradable sensorialmente que el que tiene una mayor

proporción de omaga-3, lo que indica que este aderezo en particular resulta la mejor opción para incorporar ácidos grasos omega-6 y omega-3 en la dieta de las personas.

REFERENCIAS

Carrero, J., Salmeron, L., Ramos, G., López, H., Ros, D. (2006) Estudio de las repercusiones clínicas y analíticas de una intervención nutricional en pacientes no hospitalizados con claudicación intermitente. Estudio aleatorio controlado. *Angiologia* (58), 19-30.

Simopoulos, A. (2004) Omega-6/Omega-3 essential fatty acid ratio and chronic diseases. *Food reviews international* , **20** (1), 77-90.

StatSoft, Inc. 2004. STATISTICA (data analysis software system), version 7. www.statsoft.com.

WHO. (2008). *World Health Organization*. Consultado el 01 de Junio de 2009, de http://www.who.int/cardiovascular_diseases/en/index.html

Comparación de Filtros Específicos y no Específicos en el Análisis de Matrices de Expresión Génica

María del Carmen Romero¹, Julio Alejandro Di Rienzo²

¹Universidad Nacional del Centro de la Provincia de Buenos Aires mariadelc.romero@gmail.com

²Estadística y Biometría. Fac. Cs. Agropecuarias. Universidad Nacional de Córdoba

Resumen

En este trabajo se realizó la comparación entre los *rankings* de atributos generados por filtros específicos y no específicos mediante una medida que se basa en la proporción de atributos “correctamente ubicados” (Romero y Di Rienzo, 2008). Entre los filtros específicos se consideraron el estadístico F de análisis de varianza, RELIEF-F (Kononenko, 1994), RELIEF-Modificado (Romero y Di Rienzo, 2007), Vecinos (Romero y Di Rienzo, 2005), CHI2 (Liu y Setiono, 1995) y SOAP (Ruíz Sánchez, 2006) y entre los no específicos se tuvieron en cuenta los implementados en el paquete *genefilter* de la biblioteca del proyecto *BioConductor*. La comparación se realizó para datos simulados de micromatrices de ADN y se trabajó con distintos escenarios de simulación que varían en la cantidad de atributos, clases, repeticiones y en el porcentaje de genes diferenciales.

Palabras clave: micromatrices de ADN, filtros específicos, filtros no específicos, ranking de atributos.

Introducción

Los experimentos de genómica funcional basados en micromatrices de ADN, generan, no importando la plataforma, matrices de expresión génica (MEG). Éstas permiten el monitoreo de los niveles de expresión de miles de genes simultáneamente (Dudoit *et al.*, 2002). Las filas de estas matrices corresponden con los genes evaluados y las columnas a las micromatrices utilizadas para monitorear su nivel de expresión. Estas micromatrices son dispositivos que permiten obtener una medida relativa de la expresión de un conjunto de genes en una muestra biológica. Cada micromatriz corresponde, en un experimento típico, a una réplica biológica y están agrupadas por la condición experimental (tratamiento) que les dan origen. Por ejemplo, en un experimento puede compararse la actividad génica en tejidos de plantas control y en tejidos de plantas sometidas a estrés abiótico. En este ejemplo hay dos condiciones experimentales representadas por los tipos de plantas y habrá un conjunto de réplicas biológicas (planta-micromatriz) para cada una de ellas. Usualmente, la cantidad de genes evaluados es comparativamente superior a la cantidad de micromatrices (se corresponde con la cantidad de tratamientos o materiales evaluados multiplicado por el número de repeticiones biológicas disponibles para cada tratamiento). El principal objetivo del análisis de la MEG es establecer cuáles son los genes que responden de manera diferencial bajo los distintos tratamientos.

Para usar una terminología común en la minería de datos, nos referiremos a las filas (genes) y columnas (micromatrices) de la MEG como *atributos* e *instancias* respectivamente y a los tratamientos considerados: *clases*.

La aplicación de filtros para la reducción del número de atributos en problemas donde su número es grande, es una práctica frecuente en minería de datos. Existen muchos tipos de filtros que pueden clasificarse en específicos y no específicos (Gentleman *et al.*, 2005). Los filtros específicos (*covariate-based filters*) seleccionan atributos que optimizan un clasificador. Estos filtros explotan la variación de los atributos dentro y entre clases para definir un *score* para cada atributo de acuerdo a su rendimiento en el clasificador. Estos filtros dependen de la existencia de una covariable en los datos que identifica a las clases. Cualquier prueba estadística (como por ejemplo, prueba-t o Anova) puede ser implementada como un filtro. El filtrado no específico no supone la existencia de una covariable que identifica las clases. En este caso, la función filtro tiene parámetros específicos del contexto para definir su comportamiento. Por ejemplo, un gen se considera potencialmente interesante si tiene niveles de expresión que sean lo suficientemente variables para sustentar la hipótesis de expresión diferencial.

Si bien, la práctica más comúnmente utilizada es aplicar un filtrado no específico como pre-procesamiento de las MEGs, existe poca información sobre el comportamiento comparativo de los filtros en esa etapa del análisis.

En estudios preliminares, Romero y Di Rienzo (2008) propusieron una medida para evaluar y comparar filtros que generan un *ranking* de atributos a partir de datos simulados. El rendimiento (capacidad que tiene el filtro para seleccionar los atributos diferenciales) se basa en la proporción de atributos "correctamente ubicados" por el filtro tomando como referencia el orden esperado según el algoritmo de generación de datos.

En este trabajo se compararon filtros específicos con filtros no específicos en el contexto del análisis de matrices de expresión génica. Esta comparación se realizó teniendo en cuenta el rendimiento y la eficiencia computacional.

Materiales y Métodos

Entre los filtros específicos, se consideraron aquellos que generan un *ranking* (ordenamiento) de atributos según su capacidad para la distinción de clases ya que se caracterizan por ser muy eficientes computacional y algorítmicamente. La habilidad para distinguir clases se sintetiza en un índice que llamaremos *relevancia* y que cada método calcula de forma diferente. En todos los casos mayor relevancia implica mayor distinción entre las clases de acuerdo al atributo evaluado. Se consideraron: Estadístico F de Análisis de Varianza (F), RELIEF-F (Kononenko, 1994), RELIEF-Modificado (Romero y Di Rienzo, 2007), Vecinos (Romero y Di Rienzo, 2005), CHI2 (Liu y Setiono, 1995) y SOAP (Ruíz Sánchez, 2006).

En cuanto a los filtros no específicos, se seleccionaron algunos de los filtros implementados en el paquete *genefilter* (*BioConductor*) (Gentleman *et al.*, 2009). La implementación de cada uno de estos filtros devuelve un valor lógico para cada uno de los atributos: el valor TRUE indica que el atributo "pasa" el filtro y el valor FALSE que no lo pasa. Debido a que el objetivo es realizar una comparación de *rankings* de atributos, los filtros ya implementados fueron modificados para obtener dicho *ranking*. Esta modificación se

realizó respetando la filosofía que sigue el algoritmo y haciendo uso de sus cálculos intermedios.

El detalle del funcionamiento de los filtros no específicos se encuentra en la documentación del paquete *genefilter*. El cálculo de la relevancia de cada atributo para estos filtros se detalla a continuación:

kOverA: Se calcula según dos criterios: la suma de todos sus elementos (kOverASum) y considerando la suma de todos sus elementos según el valor absoluto (kOverAAbs).

cv: La relevancia está dada por el coeficiente de variación (CV).

maxA: La relevancia de cada atributo se calcula como el valor máximo de cada uno de ellos (MaxA).

pOverA: La relevancia de los atributos se calcula según los siguientes percentiles: 99 (pOverA99), 95 (pOverA95), 90 (pOverA90), 75 (pOverA75) y 50 (pOverA50).

gapFilter: La relevancia de los atributos se calcula según dos criterios: Recorrido intercuartílico (GAP-IQR) y Mayor diferencia entre dos valores consecutivos (GAP-MAX).

Al igual que en el caso de los filtros específicos, la mayor relevancia se asocia con mayor distinción entre las clases por parte del atributo.

La comparación de los filtros se realizó a partir del “rendimiento”, medida propuesta por Romero y Di Rienzo (2008) para evaluar filtros que generan un *ranking* de atributos. Teniendo en cuenta que a cada atributo se le puede asignar un número de orden, ya sea en base a los efectos asignados durante la simulación, en adelante ordenamiento natural, o en base a la “relevancia” computada por cada uno de los filtros, es posible apreciar el desempeño de un filtro comparando las posiciones en la lista ordenada por relevancias con las posiciones en el ordenamiento natural dado por la magnitud de los “efectos de clase”. El “rendimiento” contabiliza la cantidad (o proporción) de atributos “bien ubicados” (con respecto al ordenamiento natural), mediante el siguiente estadístico:

$$\text{Rendimiento} = \left(\frac{\sum_{i=1}^{\text{CantAtribDif}} \text{Acierto}_i}{\text{CantAtribDif}} \right) \times 100\%$$

$$\text{donde } \text{Acierto}_i = 1 - \frac{|\text{PosActual}_i - \text{PosNatural}_i|}{\text{CantidadAtributos}}$$

(*PosActual_i* hace referencia al orden o posición que ocupa el *i*-ésimo atributo según el filtro aplicado y *PosNatural_i* al orden o posición que ocupa dicho atributo en el ordenamiento generado por los “efectos de clase”). Mayor rendimiento indica mayor similitud entre el *ranking* generado por el filtro y el natural.

Resultados y Conclusiones

La comparación se realizó generando matrices de expresión génica simuladas a partir de datos de un experimento *self-self* en las que se incluyó una proporción conocida de genes diferenciales con distinto grado de diferenciación. Los escenarios de simulación con los cuales se trabajó variaron en la cantidad de genes (1000 y 5000), clases o condiciones experimentales (2 y 5), repeticiones o réplicas biológicas (2 y 5) y en el porcentaje de genes diferenciales (1% y 10%) y se realizaron 100 ejecuciones para cada escenario. En el caso

del filtro CHI2 sólo se efectuaron 100 simulaciones para los escenarios con 2 grupos, 2 réplicas y 1000 genes, para las demás situaciones se trabajó con 10 repeticiones debido al tiempo computacional que demanda este método. La comparación de las distintas estrategias de filtrado se realizó a través del rendimiento promedio y de la eficiencia computacional (tiempo promedio en segundos) – Tabla 1 –.

Tabla 1: Rendimiento promedio y tiempo promedio (en segundos) para la prueba F de análisis de varianza (F) y los filtros específicos y no específicos para escenarios con 2 y 5 grupos, 2 y 5 réplicas, 1000 y 5000 atributos y 1 y 10% de atributos diferenciales

| | | Rendimiento promedio | Tiempo promedio (en segundos) |
|-------------------------------|-------------------|-----------------------------|--------------------------------------|
| Filtros específicos | RELIEF-Modificado | 82.5113 | 2.53 |
| | SOAP | 86.9794 | 2.13 |
| | CHI2 (*) | 89.5314 | 193 |
| | Vecinos | 90.5551 | 3.74 |
| | F | 91.6358 | <0.01 |
| | RELIEF-F | 91.9078 | 4.79 |
| Filtros no específicos | pOverA50 | 52.0785 | 0.57 |
| | kOverASum | 52.4773 | 0.03 |
| | GAP-MAX | 71.6234 | 0.43 |
| | pOverA75 | 71.8576 | 0.57 |
| | CV | 72.3168 | 0.37 |
| | pOverA90 | 75.2527 | 0.57 |
| | pOverA95 | 75.5928 | 0.57 |
| | pOverA99 | 76.0940 | 0.57 |
| | MaxA | 76.1806 | 0.03 |
| | kOverAAbs | 78.1615 | 0.03 |
| | GAP-RIQ | 84.0768 | 0.43 |

(*) Se realizaron 100 simulaciones en todos los casos, salvo en * (filtro CHI2) para el cual sólo se ejecutaron 100 simulaciones para los casos con 2 grupos, 2 réplicas y 100 genes y 10 simulaciones para los demás escenarios debido al tiempo computacional demandado.

- ✓ Los filtros específicos presentan rendimientos entre el 82% y el 92%, correspondiendo el menor al filtro RELIEF-Modificado y el mayor a RELIEF-F. En cuanto al tiempo promedio de ejecución (en segundos), el filtro CHI2 requiere un tiempo de ejecución considerablemente superior al resto, en el mejor de los casos (con respecto al filtro RELIEF-F) es aproximadamente un 4000% más demandante computacionalmente.
- ✓ Los filtros no específicos considerados presentan un rendimiento entre el 52% y el 84%, correspondiendo el menor al filtro pOverA50 y el mayor a GAP-RIQ. Todos los filtros, excepto POverA50 y kOverASum presentan rendimientos superiores al 70% pero inferiores a 84%.
- ✓ En general, como es de esperar, los filtros específicos poseen un rendimiento superior al de los filtros no específicos, pero demandan un tiempo de ejecución mayor, aunque, excepto por CHI2 no constituye un factor limitante para su utilización.
- ✓ A partir de los resultados obtenidos, se recomienda el uso de filtros específicos y, dentro de ellos, el uso del filtro RELIEF-F. Este filtro calcula la relevancia en base a la

distancia entre instancias definida en el espacio de todos los atributos. Por ello, la valoración que realiza de cada atributo, la hace en un contexto multivariado a diferencia de los otros métodos. Esta valoración, contextualizada por todos los atributos considerados, produce un ranking más robusto que se refleja en el mayor rendimiento observado.

REFERENCIAS

- Dudoit, S.; Fridlyand, J. y Speed, T. (2002). "Comparison of discrimination methods for the classification of tumors using gene expression data". *Journal of the American Statistical Association*, 97(457): 77-87.
- Gentleman, R.; Carey, V. J.; Huber, W.; Irizarry, R. y Dudoit, S. (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Statistics for Biology and Health. Ed. Springer Science+Business Media, Inc.
- Gentleman, R.; Carey, V.; Huber, W. y Hahne, F. (2009). Manual de Referencia del package Genefilter. <http://www.bioconductor.org/packages/2.5/bioc/manuals/genefilter/man/genefilter.pdf>
- Kononenko, I.; Smic, E. y Robnik-Sikonja, M. (1997). Overcoming the myopia of induction learning algorithms with RELIEFF. *Applied Intelligence* 7(1), 39-55.
- Liu, H. y Setiono, R. (1995). "Chi2: Feature selection and discretization of numeric attributes". *7th IEEE International Conference on Tools with Artificial Intelligence*, 388-391.
- Romero, M. del C. y Di Rienzo, J. A. (2005). "Análisis de perfiles genéticos: Selección de atributos". *IV Congreso Latinoamericano de Biología Matemática*. Tandil, Argentina. 6 al 11 de noviembre de 2005.
- Romero, M. del C. y Di Rienzo, J. A. (2007). "Comparación de distintos métodos de filtrado: Una aplicación a micromatrices de ADN". *XII Reunión Científica del Grupo Argentino de Biometría y I Encuentro Argentino – Chileno de Biometría*. San Martín de los Andes, Argentina. 10 al 12 de octubre de 2007.
- Romero, M. del C. y Di Rienzo, J. A. (2008). "Comparación de rankings de atributos". *XIII Reunión Científica del Grupo Argentino de Biometría*. Tucumán, Argentina. 17 al 19 de septiembre de 2008.
- Ruiz Sánchez, R. (2006). "Heurísticas de selección de atributos para datos de gran dimensionalidad". Memoria de Tesis Doctoral para optar al grado de Doctor en Informática por la Universidad de Sevilla. Sevilla, España.

Utilización del Análisis de Cluster con Variables Mixtas en la Identificación Taxonómica de Ácaros Depredadores (acarí phytoseiidae) de Cuba

Dunia Chávez Esponda¹, Ileana Miranda Cabrera², Mayra Ramos Lima³

¹Universidad Agraria de La Habana (UNAH). Cuba. dunia@isch.edu.cu

²Centro Nacional de Sanidad Agropecuaria (CENSA). Cuba. ileanam@censa.edu.cu

³Instituto de Investigaciones de Sanidad Vegetal (INISAV). Cuba. mayramos@inisav.cu

Resumen

El diagnóstico de ácaros depredadores, es la base para establecer un adecuado control biológico. Esta identificación taxonómica es posible obtenerla a través de métodos multivariados. El Análisis de Cluster es un método de clasificación que se utiliza para establecer grupos de individuos y está asociado a la taxonomía numérica de especies. Parte de un índice de similitud o distancia el cual se selecciona de acuerdo a las variables medidas. Cuando hay variables mixtas (cuantitativas, cualitativas y binarias) se propone usar el coeficiente de Gower, sin embargo en Cuba, su uso es muy escaso. En este trabajo se utilizó el Análisis de Cluster a partir del coeficiente de similaridad de Gower, con el objetivo de estudiar la taxonomía de los ácaros fitoseidos cubanos. Este método permitió diferenciar especies que poseían características muy similares, lo cual es de gran importancia para los taxónomos identificar de una manera confiable y rápida una especie.

Palabras clave: análisis de cluster, similaridad de Gower, taxonomía, ácaros fitoseidos.

Introducción



Los ácaros de la familia Phytoseiidae se reconocen como excelentes biorreguladores de ácaros fitoparásitos y de algunos insectos. El estudio de su taxonomía se ha convertido en un tema de interés en muchas investigaciones actuales. La correcta identificación taxonómica de estos depredadores posibilita un adecuado control biológico.

El Análisis de Cluster es el método esencial para hacer estas clasificaciones. Este algoritmo depende de la correcta selección del coeficiente de similaridad o distancia y del método de agrupamiento a utilizar. Aunque algunos autores han tratado de implementar distancia que permitan un tratamiento para unificar la taxonomía convencional con la molecular, hasta hoy la Estadística no ha ofrecido una solución algorítmica para obtener los resultados más confiables y los biólogos y taxónomos desconocen las ventajas y desventajas de una u otra métrica. En la literatura se reporta que en presencia de variables mixtas (cuantitativas, cualitativas y binarias) se propone usar el coeficiente de Gower (GOWER, 1971), sin embargo en la esfera agropecuaria de Cuba, su uso es nulo.

En este trabajo se empleó el Análisis de Cluster a partir del coeficiente de similaridad de Gower, con el objetivo de clasificar taxonómicamente los ácaros cubanos de la familia Phytoseiidae.

Desarrollo

El coeficiente de Gower tiene por expresión:

$$d_{ij}^2 = 1 - s_{ij}$$

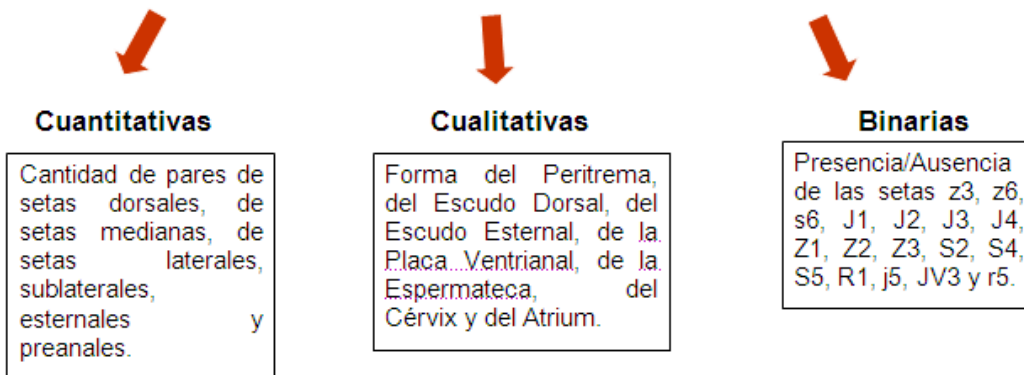
$$s_{ij} = \frac{\sum_{h=1}^{p1} \left(1 - \frac{|x_{ih} - x_{jh}|}{G_h} \right) + a + \alpha}{p1 + (p2 - d) + p3}$$

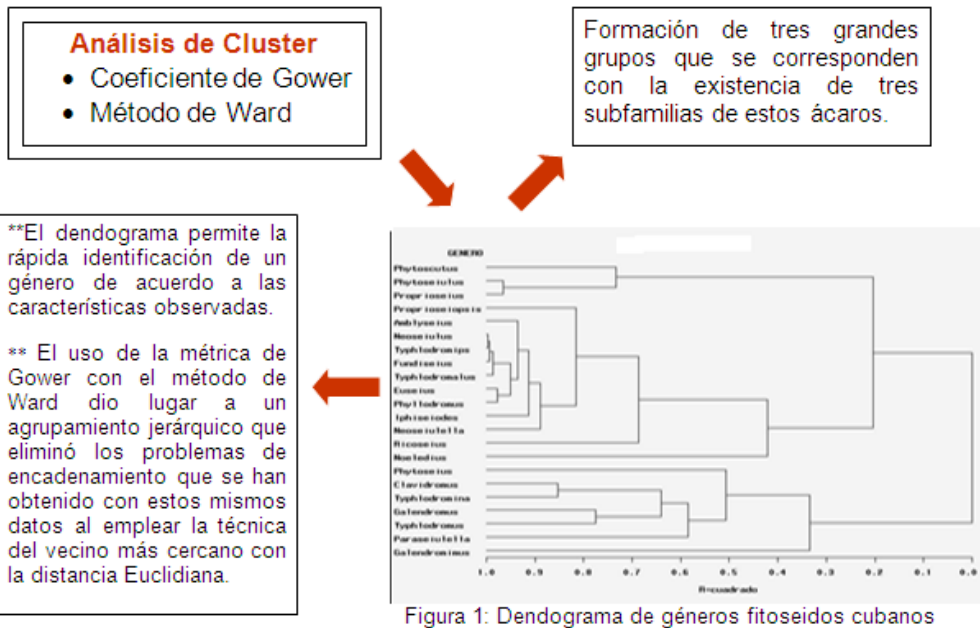
p1: número de variables cuantitativas,
a: número de coincidencias en 1
d: número de coincidencias en 0 de las p2 variables binarias
α: es el número de coincidencias para las p3 variables cualitativas.
G_h es el rango de la h-ésima variable cuantitativa.
Rango= X máximo - X mínimo.

Ventajas del uso de este coeficiente para la taxonomía de las especies

- Las características morfológicas y moleculares que se miden en las especies involucran variables de diferente naturaleza (cuantitativa, cualitativa, binaria); el coeficiente de similitud de Gower (GOWER, 1971) es apto para calcular similitudes cuando se tienen tales mezclas de variables.
- Es posible trabajar con bases de datos en las que faltan observaciones de algunas variables, sin prescindir de todo el vector que representa a la unidad muestral ni usar ningún método de imputación.
- Mediante el uso de este coeficiente es posible ponderar las variables de manera diferencial, dependiendo del papel que se quiera que cada una juegue en la ordenación.

En el estudio de la taxonomía de los ácaros fitoseidos están presentes caracteres mixtos (cuantitativos, cualitativos y binarios) los cuales constituyen estructuras importantes en la identificación de los géneros y especies. En este trabajo se tuvieron en cuenta las siguientes variables:





Finalmente se utilizó el Análisis Discriminante para validar los grupos formados. Teniendo en cuenta la presencia de variables mixtas, se utilizó el procedimiento de selección de variables paso a paso. Se realizó la estimación del error de clasificación por el método de validación cruzada en el cual se observó en la matriz de clasificación un 93% de buena clasificación. Por otro lado, se obtuvo el valor del coeficiente de correlación cofenética ($rc=0.891$) el cual permite medir la calidad de la clasificación obtenida (dendograma). Los valores obtenidos permitieron concluir que la agrupación fue aceptable.

Tal y como señaló GOWER (1967), el método del vecino más cercano no ofrece grupos claramente separados porque tiende a una agrupación longitudinal. Luego, utilizar diferentes medidas de distancia y distintos criterios de enlace sigue siendo la manera más recomendada para obtener resultados consistentes y agrupamientos naturales.

Conclusiones

- Se obtuvieron tres grupos de géneros en los cuales se agrupan varias especies con características similares. Esto se corresponde con la presencia de tres subfamilias en la familia Phytoseiidae.
- El dendograma obtenido posibilita una rápida identificación de los géneros fitoseidos cubanos, lo cual favorece establecer una adecuada estrategia de manejo integrado de plagas.
- Se recomienda utilizar la distancia de Gower y el método de Ward para agrupamiento jerárquico a partir de caracteres mixtos.

Un Modelo SARIMA para Viento en la Estación Nucleoeléctrica de Laguna Verde

Uziel González González¹, Alejandro Jiménez Lagunes², Sergio F. Juárez Cerrillo³

¹Facultad de Matemáticas, UNAM.

²Doctorado de Geografía, UNAM.

³Facultad de Estadística e Informática, Universidad Veracruzana.

1. Introducción

Una prioridad de la Central Núcleo-eléctrica Laguna Verde (CNLV), central ubicada en la zona centro-este del estado de Veracruz (véase la Figura 1), es contar con un modelo para la simulación del campo de viento. Esto ante la necesidad de validar un modelo de dispersión de contaminantes que mejor se adecue a las características geográficas de la zona de Laguna Verde. Dicho modelo de dispersión tendría la finalidad de anticipar los alcances geográficos de los contaminantes expulsados a la atmósfera durante una posible expulsión de contaminantes a la atmósfera. En la CNLV ya se cuenta con un modelo de viento, un modelo Gaussiano recomendado por el Organismo Internacional de Energía Atómica, sin embargo, estudios han determinado que este modelo no es adecuado debido a las complejas condiciones climáticas y orográficas de la región aledaña a la CNLV (terreno accidentado con altitudes que varían de 0 m a 1000 msnm en un radio de ~20 km); Cervantes (1987), Morales et al. (1989) Contreras (1989), Tejeda et al. (1994), Tejeda et al. (1989), Hernández (1999), y Ramos (2000). Además, en estos estudios se demuestra que el desarrollo de un modelo dinámico tampoco es posible. Es por ésta razón que estos trabajos tienen un enfoque estadístico.

En este trabajo damos el primer paso para el desarrollo de un modelo para un campo continuo a partir de un conjunto de datos discretos. Es decir, proponemos modelos tentativos para los datos de viento registrados en la estación meteorológica de la CNLV con el objetivo de sentar las bases para un modelo espacio-temporal para la capa de viento en la zona de la CNLV.



Figura 1. Vistas de la Central Núcleo-eléctrica de Laguna Verde.

2. Datos

Los datos de viento que se analizan fueron registrados en la CNLV de manera horaria a partir del 1 de enero de 1980 hasta 31 de diciembre del 2007. Esto resulta en cuatro series de tiempo con 245, 448 observaciones: la velocidad a 10 metros y 60 metros (WS10 y WS60), y la dirección a 10 y 60 metros (WD10 y WD60). La estructura de la base de datos se muestra en la Tabla 1 para las primeras 10 observaciones.

| WS10 | WS60 | WD10 | WD60 | HORA | DIA | MES | AÑO |
|------|------|-------|-------|------|-----|-----|------|
| 5.9 | 6.6 | 315 | 337.5 | 1 | 1 | 1 | 1980 |
| 4.9 | 5.6 | 315 | 337.5 | 2 | 1 | 1 | 1980 |
| 4.3 | 5 | 315 | 337.5 | 3 | 1 | 1 | 1980 |
| 4.6 | 5.2 | 315 | 337.5 | 4 | 1 | 1 | 1980 |
| 5.6 | 5.8 | 337.5 | 360 | 5 | 1 | 1 | 1980 |
| 6.3 | 7.7 | 315 | 315 | 6 | 1 | 1 | 1980 |
| 5.7 | 4.7 | 315 | 337.5 | 7 | 1 | 1 | 1980 |
| 5.2 | 4.9 | 337.5 | 337.5 | 8 | 1 | 1 | 1980 |
| 6.3 | 6.3 | 337.5 | 337.5 | 9 | 1 | 1 | 1980 |
| 4.7 | 4.5 | 337.5 | 337.5 | 10 | 1 | 1 | 1980 |

Tabla 1. Datos de viento de la estación de Laguna Verde.

Para evitar el tratamiento especializado que requieren las variables circulares (el viento es una variable circular vectorial), en este artículo decidimos trabajar con las componentes rectangulares del viento en lugar de las componentes circulares, es decir, analizamos a $X_{10} = WS_{10} \cos(\theta_{WD10})$, $Y_{10} = WS_{10} \sin(\theta_{WD10})$, $X_{60} = WS_{60} \cos(\theta_{WD60})$, y $Y_{60} = WS_{60} \sin(\theta_{WD60})$.

3. Modelos SARIMA y Pronóstico

Después de realizar el proceso de modelación estadística siguiendo la metodología de Box-Jenkins así como el diagnóstico para evaluar la calidad del ajuste, se seleccionaron los siguientes modelos SARIMA para cada una de las series de tiempo analizadas.

| Serie de tiempo | Modelo |
|-----------------|-----------------------------|
| X10 | SARIMA (1,1,1)x(2,1,3) S=24 |
| Y10 | SARIMA (3,1,1)x(2,1,1) S=24 |
| X60 | SARIMA (1,1,1)x(2,1,2) S=24 |
| Y60 | SARIMA (3,1,1)x(1,1,1) S=24 |

Tabla 2. Modelos ajustados.

El pronóstico para cada una de las cuatro series de tiempo se presenta en la Figura 2.

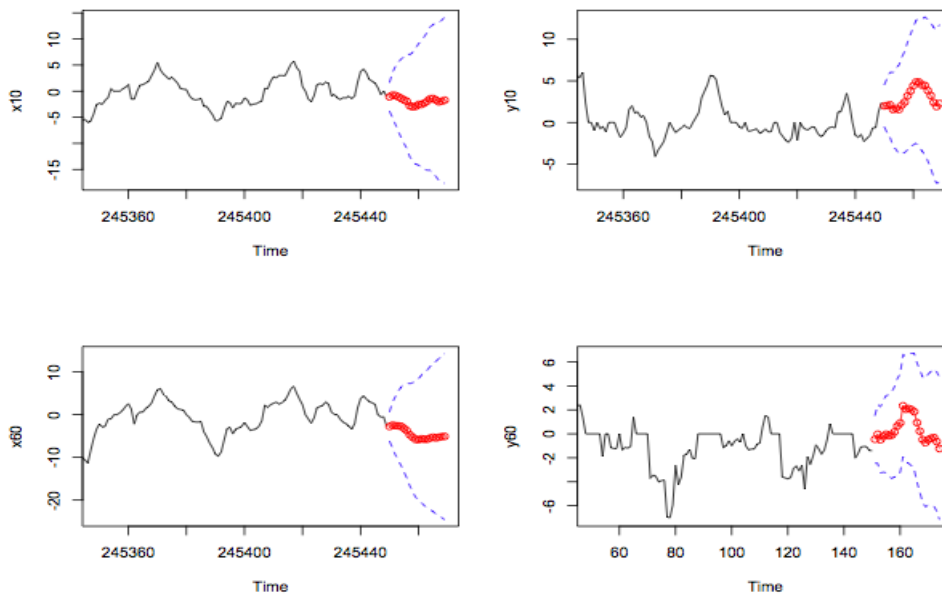


Figura 2. Pronóstico para las series de tiempo.

4. Investigación Futura

Como se mencionó anteriormente, se han realizado diversos estudios encaminados a conocer los patrones de circulación en los alrededores de la CNLV. Estos estudios se han hecho con datos de cuatro estaciones circunvecinas de la CNLV, las cuales dejaron de funcionar en 1990. Posteriormente, debido a la preocupación de la CNLV por continuar con los estudios para determinar los patrones de circulación en la zona, en 1998 se pusieron en operación tres estaciones anemométricas en las localidades de Úrsulo Galván, El Diamante y Santa Ana (Figura 3). Con los datos generados en estas estaciones y con el objetivo de equipar a la CNLV, Hernández (1999) usó el modelo de mesoescala MM5 para simular el campo de viento. En el estudio de Ramos (2000), de carácter estadístico, se postularon modelos de regresión lineal múltiple que explotan la relación entre los datos de viento de las nuevas estaciones y los datos de viento en Laguna Verde. Estos modelos permiten predecir los datos de viento (dirección y magnitud) en las tres nuevas estaciones cuando por alguna razón éstas no cuentan con datos observados. No obstante y a pesar de todos los esfuerzos realizados con estos valiosos estudios, en la actualidad, la CNLV no cuenta con un modelo que permita simular el campo de viento. Por lo tanto, el objetivo de este trabajo es sentar las bases para un modelo estadístico que permita simular el campo de viento en la zona y sus alrededores.

El modelo a desarrollar es un modelo estadístico espacio-temporal de interpolación horizontal que permita simular el campo de viento en superficie. Dicho modelo proporcionará a la CNLV datos de viento en un malla regular que para efectos de este

estudio contará de 256 puntos (la malla es de 6' de latitud x 6' longitud) con lo que se cubre aproximadamente una zona de 80 km alrededor de Laguna Verde (Figura 3) la cual incluye la zona del Plan de Emergencia Radiológica Externa PERE (cuadro rojo en la Figura 1). El modelo horizontal incluye un factor que permite parametrizar las características del terreno en la estación anemométrica y el punto de malla.

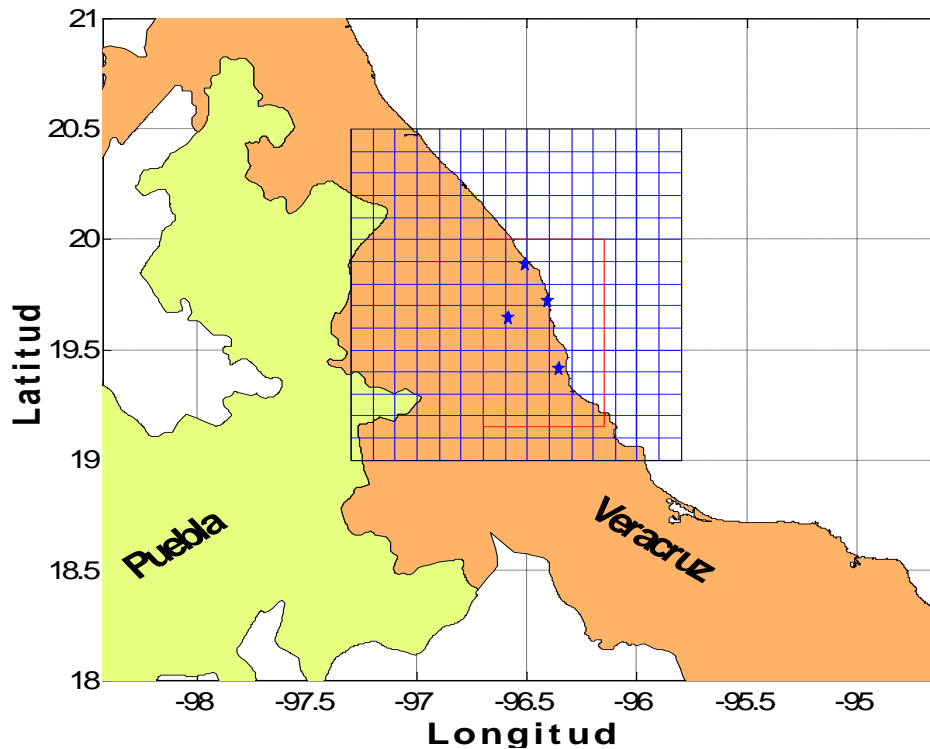


Figura 3. Zona de estudio y la malla (256 nodos) en donde el modelo propuesto deberá proporcionar datos de viento (magnitud y dirección). Las estaciones anemométricas en azul (asteriscos). El cuadro en rojo hace referencia a la zona de estudio primordial.

Referencias

- Cervantes, J. (1987). Relaciones entre la dirección del viento a 850 mb y la dirección del viento en superficie en la zona de Laguna Verde, Veracruz. Tesis de Licenciatura en Ciencias Atmosféricas, Universidad Veracruzana, Xalapa. pp 9-16.
- Contreras, A. D. (1989). Los vientos locales en la costa centro-sur de Veracruz y su relación estadística a partir de la circulación de 500 mb. Tesis de Licenciatura en Ciencias Atmosféricas, Universidad Veracruzana. pp 7-13.

Hernández, R. V. (1999). Modelación preliminar del campo de viento alrededor de la central nucleoelectrica Laguna Verde. Tesis de licenciatura en Ciencias Atmosféricas, Universidad Veracruzana, Xalapa. Pp 1-72.

Morales, A. T., Tejeda, M. A. y Cervantes, P. J. (1989). The surface wind direction as a function of the wind at 850 mb in Laguna Verde zone, Veracruz, Mexico. *Atmósfera*, Vol. 2: 181-187.

Ramos, G. E., (2000). Comportamiento de la dirección y magnitud del viento en Laguna Verde y sus alrededores". Tesis de licenciatura en Ciencias Atmosféricas, Universidad Veracruzana, Xalapa. Pp 1-45.

Tejeda, A., Ojeda, M. M. and Contreras, A. D. (1994). Local Winds South of Sierra de Misantla (México), and their statistical relationships with 500 mb circulation. *Geofísica Internacional*, Vol. 33. Num 2, pp 303-309.

Tejeda, A., Morales, T., and Cervantes, J. (1989). The surface wind direction as a function of the wind at 850 mb in Laguna Verde zone, Veracruz, México. *Atmósfera* 2:181-187.

Análisis Espectral Singular de Series de Tiempo

Alejandro Ramírez Núñez¹, Sergio Francisco Juárez Cerrillo²

¹Facultad de Matemáticas. Universidad Autónoma de Guerrero. alex_ramirez_1987@hotmail.com

²Facultad de Estadística e Informática, Universidad Veracruzana. sejuarez@uv.mx

Resumen

El Análisis Espectral Singular (AES) es una técnica de desarrollo reciente para analizar series de tiempo. En esencia, el AES es un análisis de componentes principales en el dominio del tiempo de una serie de tiempo. El objetivo del AES es extraer información de la dinámica de una serie de tiempo y resulta una herramienta poderosa para analizar series de tiempo cortas y con mucho ruido. Es por esto que el AES se ha aplicado ampliamente en geofísica, climatología, meteorología y recientemente en econometría y finanzas. En este artículo presentamos brevemente el AES y lo ilustramos con el análisis de la serie de temperaturas máximas diarias registradas en la estación Actopan estado de Veracruz de 1991 al 2002. La implementación del AES la hacemos con el programa S-Plus.

Palabras Clave: Análisis de Componentes Principales, Análisis de Series de Tiempo en el Dominio del Tiempo, Descomposición en Valores Singulares, Eigenanálisis.

1. Introducción al Análisis Espectral Singular

El Análisis Espectral Singular (AES) es una técnica no paramétrica de análisis de series de tiempo que consta de dos etapas divididas en los siguientes pasos:

- 1° Construcción de la matriz trayectoria
- 2° Descomposición en valores singulares
- 3° Agrupación
- 4° Promediación diagonal

Consideremos a la serie de tiempo real $Z = (z_0, \dots, z_{N-1})$ de longitud $N > 2$.

Primera etapa: Descomposición.

1° paso: Transformación.

En el primer paso se mapea la serie de tiempo original a una secuencia de vectores del siguiente modo. Sea L un número entero tal que $1 < L < N$, llamaremos a L la longitud de ventana. Se forman $K = N - L + 1$ vectores de retraso

$$\mathbf{X}_i = (z_{i-1}, \dots, z_{i+L-2})^T, \quad 1 \leq i \leq K.$$

Notemos que estos vectores son de dimensión L . Con los vectores de retraso se construye la *matriz de trayectoria* \mathbf{X} de la serie Z

$$\mathbf{X} = [X_1, X_2, \dots, X_K].$$

Por lo que los vectores de retraso son las columnas de \mathbf{X} . Esto es, la matriz trayectoria es

$$\mathbf{X} = \begin{pmatrix} z_0 & z_1 & z_2 & \cdots & z_{K-1} \\ z_1 & z_2 & z_3 & \cdots & z_K \\ z_2 & z_3 & z_4 & \cdots & z_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{L-1} & z_L & z_{L+1} & \cdots & z_{N-1} \end{pmatrix}. \quad (1)$$

Es claro que $x_{ij} = z_{i+j-2}$, por lo que la matriz \mathbf{X} tiene los elementos iguales en la diagonal $i + j = \text{constante}$. La matriz de trayectoria es conocida, por su estructura, como matriz de Hankel. Si N y L se fijan, entonces hay una correspondencia uno-a-uno entre las matrices de trayectoria y las serie de tiempo.

2° paso: Descomposición en Valores Singulares

En este paso se calcula la descomposición en valores singulares (DVS) de \mathbf{X}

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T,$$

donde las columnas de \mathbf{U} , $\mathbf{U}_1, \dots, \mathbf{U}_L$, se conocen como los vectores singulares izquierdos de \mathbf{X} , las columnas de \mathbf{V} , $\mathbf{V}_1, \dots, \mathbf{V}_K$, se conocen como vectores singulares derechos de \mathbf{X} y $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d)$ donde d es el rango de \mathbf{X} . Los números $\lambda_1 \leq \dots \leq \lambda_L$ se conocen como los valores singulares de \mathbf{X} . Consideremos a las matrices de rango 1 $\mathbf{X}_i = \lambda_i \mathbf{U}_i \mathbf{V}_i^T$, $i = 1, \dots, d$. Entonces la matriz de trayectoria \mathbf{X} puede escribirse como

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d \quad (2)$$

Al triple $(\lambda_i, \mathbf{U}_i, \mathbf{V}_i)$ lo llamaremos el *eigen triple* i de la DVS de \mathbf{X} .

Segunda etapa: Reconstrucción.

3° Paso: Agrupación.

Una vez que se ha obtenido la DVS de \mathbf{X} , se realiza el procedimiento de agrupación. En este paso se divide al conjunto de índices $\{1, \dots, d\}$ en m subconjuntos disjuntos $\mathbf{I}_1, \dots, \mathbf{I}_m$.

Sea $\mathbf{I} = \{i_1, \dots, i_p\}$, entonces la matriz resultante $\mathbf{X}_{\mathbf{I}}$ correspondiente al grupo \mathbf{I} se define por

$\mathbf{X}_{\mathbf{I}} = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. Entonces, de la DVS en (2) tenemos la descomposición:

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}. \quad (3)$$

Al procedimiento de elegir los conjuntos I_1, \dots, I_m se llama agrupamiento de los eigentriples.

4° Paso: Promediación Diagonal

El último paso en el AES consiste en transformar cada matriz de la descomposición agrupada (3) en una nueva serie de tiempo de longitud N .

Sea \mathbf{Y} una matriz de tamaño $L \times K$ con elementos y_{ij} , $1 \leq i \leq L$, $1 \leq j \leq K$. Sea $L^* = \min(L, K)$, $K^* = \max(L, K)$ y $N = L + K - 1$. Sea $y_{ij}^* = y_{ij}$ si $L < K$ y $y_{ij}^* = y_{ji}$ en otro caso.

La *promediación diagonal* transforma la matriz \mathbf{Y} en la serie g_0, \dots, g_{N-1} del siguiente modo:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^* & \text{para } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^* & \text{para } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^* & \text{para } K^* \leq k < N, \end{cases} \quad (4)$$

La expresión (4) corresponde a los promedios de los elementos de la matriz sobre las “diagonales” $i + j = k + 2$. Si elegimos $k = 0$ tenemos $g_0 = y_{11}$, para $k = 1$ tenemos $g_1 = (y_{12} + y_{21})/2$ y así sucesivamente. Notemos que si la matriz \mathbf{Y} es una matriz trayectoria de una serie (h_0, \dots, h_{N-1}) (es decir, si \mathbf{Y} es una matriz Hankel), entonces $g_i = h_i$ para toda i .

La promediación diagonal (4) aplicada a la matriz resultante \mathbf{X}_{I_k} produce la serie

$\tilde{z}^{(k)} = (z_0^{(k)}, \dots, z_{N-1}^{(k)})$ y la serie inicial Z se descompone en la suma de m series:

$$z_n = \sum_{k=1}^m z_0^{(k)}. \quad (5)$$

2. Ejemplo

El siguiente ejemplo ilustra el AES con la serie de temperaturas máximas diarias registradas en la estación Actopan de 1991 al 2002. La matriz de Hankel se construye con una ventana $L = 30$. La serie de tiempo se ilustra en la figura 1.

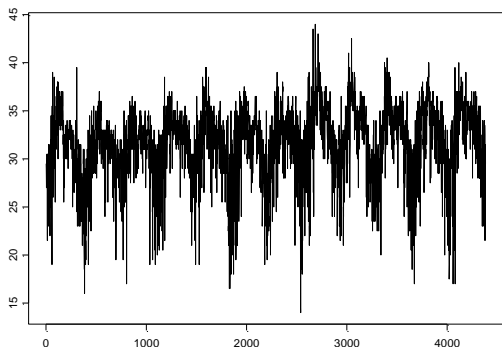


Figura 1. Temperaturas máximas diarias registradas en la estación Actopan, Veracruz de 1991 al 2002.

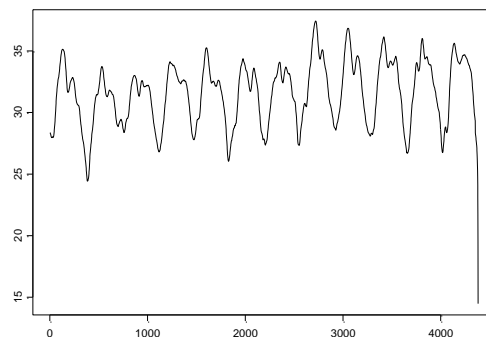


Figura 2. Primer eigentriple.

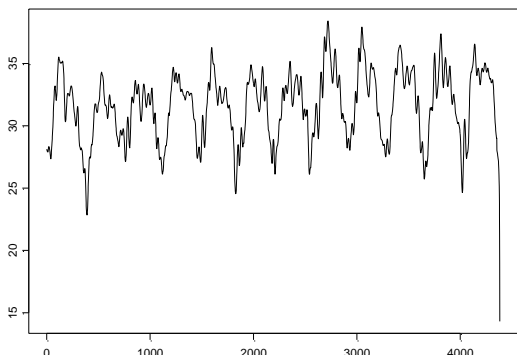


Figura 3. Primer y segundo eigentriples sumados.

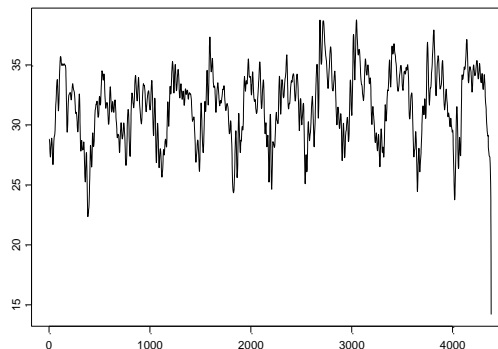


Figura 4. Primer, segundo y tercer eigentriples sumados.

Bibliografía

Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001), *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall/CRC: Boca Raton.

David H. Schoellhamer, A. (2001), *Singular Spectrum Analysis for Time Series with Missing Data*. U.S. Geological Survey, Sacramento, California.

D. Kondrashov, Y. Feliks, and M. Ghil, *Singular Spectrum Analysis with Missing Data and Oscillatory Climate Modes in Extended Nile River Record (A.D. 622-1922)*. Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California Los Angeles.

Víctor Manuel Guerrero Guzmán, A. (2003), *Análisis Estadístico de Series de Tiempo Económicas*. Thomson: México y América Central.

Análisis Estadístico de la Implementación del Laberinto Acuático de Morris en la Evaluación del Aprendizaje Espacial

Pérez Roldan E.A¹, Santiago Roque I², Juárez Aguilar E³, y Juárez Cerrillo S.F.⁴

¹ Especialidad en Métodos Estadísticos, UV. lalo_13_1970@hotmail.com

² Facultad de Bioanálisis, UV. isantiago@uv.mx

³ Instituto de Ciencias de la Salud, UV. enjuarez@uv.mx

⁴ Facultad de Estadística e Informática, UV. sejuarez@uv.mx

Resumen

Analizamos los resultados de un experimento en el cual se planteó evaluar, mediante el modelo conductual de laberinto acuático de Morris, el desarrollo del aprendizaje espacial de roedores que son entrenados para la solución de tareas. Estas tareas dependen de la adquisición de información espacial proporcionada a través de objetos. Consideramos como indicador de aprendizaje espacial la disminución en el tiempo en el que la rata localiza la plataforma escondida en varios intentos. Mediante el ajuste de un modelo de regresión no-paramétrico determinamos evidencia estadística de una disminución en el tiempo en el que las ratas encuentran la plataforma después de 24 entrenamientos (8 por día durante 3 días), y por lo tanto, de que las ratas adquieren aprendizaje espacial. También encontramos, mediante un análisis de varianza no paramétrico, evidencia significativa de que el tiempo promedio disminuye del primer al segundo día y del segundo día al tercer día. En conclusión, probamos estadísticamente que el tiempo que les toma a las ratas encontrar la plataforma después de 24 entrenamientos (8 por día en 3 días) disminuye.

Palabras Clave: Inferencia para medias con restricciones de orden, prueba de Page, regresión no paramétrica, suavizamiento de curvas con núcleos.

1. Introducción

Una de las características más importantes del ser humano es su capacidad de aprender y recordar, con ello, el puede reconocer estímulos vistos con anterioridad, asociarlos y orientarse de acuerdo con las asociaciones que realice. Esta característica es compartida con la mayoría de los animales utilizándola para la obtención de sus alimentos, conductas reproductivas, y para regresar a su nido. Es por ello que el uso de roedores como modelo para estudiar procesos neurobiológicos implicados en el aprendizaje es muy frecuente, a través de solución de tareas que dependen de información espacial disponible en el entorno. Para valorar esta conducta Morris y colaboradores en 1981 propuso el uso de un laberinto acuático, en donde el roedor nada y busca una plataforma que se encuentra oculta en la piscina, auxiliándose de claves espaciales extra-laberinto y de entrenamientos repetidos.

2. Metodología

Sea Y_{ij} el tiempo que le toma a la rata j encontrar la plataforma en el tiempo t_i . Consideramos que los tiempos dentro de las ratas están correlacionados serialmente y que

son independientes entre ratas. Supondremos que el siguiente modelo describe a los tiempos

$$Y_{ij} = m(t_i) + \varepsilon_{ij}, \quad i = 1, \dots, n, j = 1, \dots, N, \quad (1)$$

donde m es una función de regresión desconocida y los errores ε_{ij} son variables aleatorias con media cero y

$$\text{Cov}(\varepsilon_{ij}, \varepsilon_{kl}) = \begin{cases} \sigma^2 \rho(t_i - t_k), & j = l; \\ 0, & j \neq l. \end{cases}$$

La función de correlación ρ es par con $\rho(0) = 1$ y $|\rho(u)| \leq 1$ para $-1 \leq u \leq 1$. Si promediamos a los tiempos a través de las ratas

$$Y_i = \frac{1}{N} \sum_{j=1}^N Y_{ij}, \quad \varepsilon_i = \frac{1}{N} \sum_{j=1}^N \varepsilon_{ij}, \quad i = 1, \dots, n,$$

podemos escribir el siguiente modelo para los promedios de las respuestas

$$Y_i = m(t_i) + \varepsilon_i. \quad (2)$$

En este modelo los errores promediados satisfacen $\text{Cov}(\varepsilon_i, \varepsilon_k) = (\sigma^2/N)\rho(t_i - t_k)$. El problema de estimación de m en (1) se puede resolver mediante la estimación de m en (2). En este trabajo estimamos a m ajustando una curva suave que pasa a través de las medias Y_i

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^n W_{hi} Y_i,$$

donde

$$W_{hi}(x) = n \int_{s_{i-1}}^{s_i} K_h(x - u) du,$$

con $s_0 = 0$, $s_i = (t_i - t_{i+1})/2$, $i = 1, \dots, n - 1$, $s_n = 1$. Y $K_h(x) = h^{-1}K(u/h)$ donde h es el ancho de banda y K es un núcleo, es decir, K es una función real, continua, simétrica que integra 1,

$$\int_{-\infty}^{\infty} K(u) du = 1.$$

Un estimador de la función de correlación ρ es $\hat{\rho}(k) = \hat{c}(k)/\hat{c}(0)$ donde

$$\hat{c}(k) = \frac{1}{nN} \sum_{j=1}^N \sum_{i=1}^{n-k} (Y_{ij} - Y_i) (Y_{i+k,j} - Y_{i+k}).$$

Para probar si hay diferencias en la disminución en los tiempos de aprendizaje día a día, promediamos a los tiempos de cada rata por día. Los resultados se presentan en la Tabla 1. Sea X_{ij} el tiempo promedio de la rata i en el día j ($i = 1, \dots, 8, j = 1, \dots, 3$). Ahora ordenamos a los tiempos de cada rata (calculamos los rangos por hilera de la Tabla 1 y se muestran entre paréntesis), y sea R_j la suma de los rangos del día j (suma de los rangos de la columna de la Tabla 1). Supondremos lo siguiente:

- Los tiempos promedio entre ratas (bloques) son independientes.
- La variable de interés es continua.
- No hay interacción entre bloques (ratas) con los tratamientos (días).
- Las observaciones dentro del bloque (rata) se pueden ordenar por magnitud.

La hipótesis de disminución en los tiempos de día a día la probamos con

$$H_0: \mu_1 = \mu_2 = \mu_3,$$

$$H_1: \mu_1 \geq \mu_2 \geq \mu_3.$$

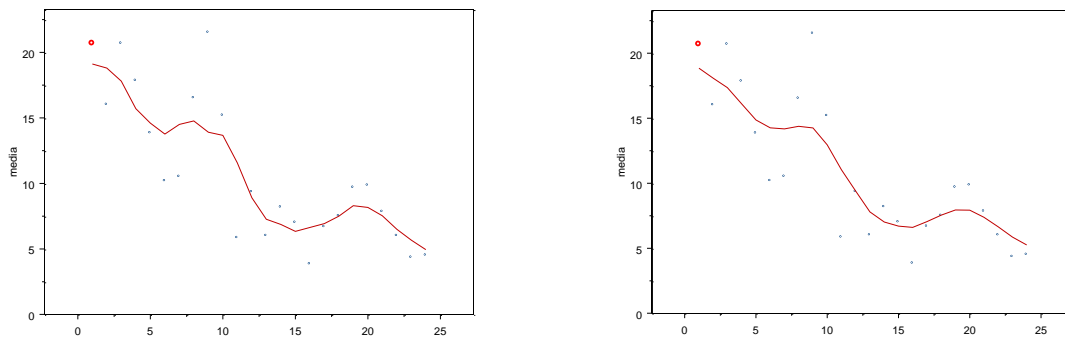
donde μ_j es la mediana de los tiempos en el día j . Para probar la hipótesis anterior utilizamos la prueba de Page de análisis de varianza de dos entradas para hipótesis alternas ordenadas, cap. 7 de Daniel (1990). El estadístico de la prueba de Page es

$$L = \sum_{j=1}^k (k + 1 - j)R_j = 3R_1 + 2R_2 + R_3.$$

La hipótesis nula se rechaza a favor de la alterna para valores grandes de L . El nivel de significancia observado lo obtenemos de la Tabla A.17 de Daniel (1990).

3. Resultados

Los datos resultantes del experimento se presentan en la Tabla 2 del Apéndice. La Figura 1 muestra las curvas de regresión no-paramétrica ajustada a los promedios (última columna de la Tabla 2) usando el núcleo rectangular, triangular, Parzen y Gaussiano; todos con un ancho de banda $h = 5$. En todas las curvas es evidente la disminución en los tiempos promedio en el que las ratas encontraron la plataforma.



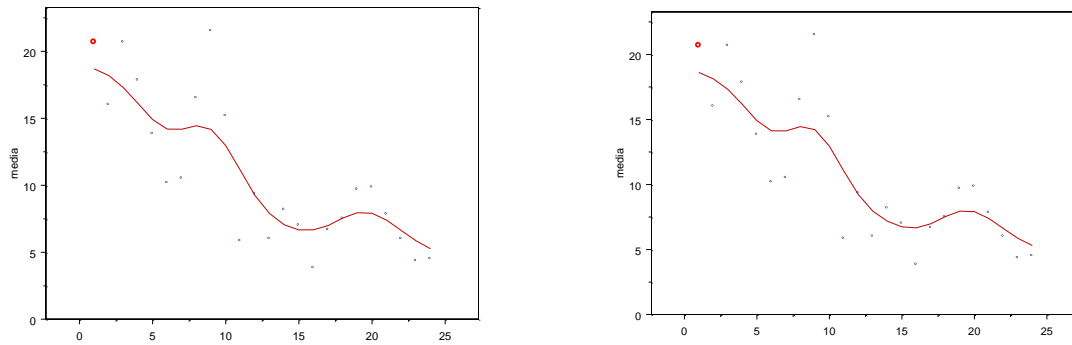


Figura 1. De izquierda a derecha y de arriba abajo, tenemos las curvas de regresión con los núcleos rectangular, triangular, de Parzen y Gaussiano.

Ahora usamos los datos de la Tabla 1

| Rata | Día | | |
|------|-----------|-----------|----------|
| | 1 | 2 | 3 |
| 1 | 27.13 (3) | 10.00 (2) | 7.38 (1) |
| 2 | 13.13 (3) | 4.50 (1) | 7.13 (2) |
| 3 | 15.13 (3) | 9.13 (2) | 6.88 (1) |
| 4 | 13.75 (3) | 5.88 (1) | 7.88 (2) |
| 5 | 7.00 (1) | 10.38 (3) | 7.38 (2) |
| 6 | 21.75 (3) | 12.50 (2) | 7.00 (1) |
| 7 | 23.13 (3) | 10.38 (2) | 8.00 (1) |
| 8 | 13.88 (3) | 9.38 (2) | 5.13 (1) |

Tabla 1. Tiempos promedio por día y los rangos entre paréntesis.

para probar la hipótesis de que las medianas diarias (medianas poblacionales) disminuyen

$$H_0: \mu_1 = \mu_2 = \mu_3,$$

$$H_1: \mu_1 \geq \mu_2 \geq \mu_3.$$

El estadístico de la prueba de Page es $L = 107$ el cual es significativo al $\alpha = 0.01$, por lo que rechazamos la hipótesis nula a favor de la hipótesis alterna. Esto es, existe evidencia significativa de que los promedios de los tiempos se reducen de día a día.

Conclusiones

Se demostró estadísticamente que la disminución del tiempo para encontrar la plataforma oculta en el laberinto acuático de Morris se relaciona con el número de entrenamiento, lo que da sustento para validar el uso del LAM como un test conductual confiable para evaluar el estudio de la memoria y aprendizaje espacial de roedores.

Referencias

- Daniel, W.W. (1990). *Applied Nonparametric Statistics*, 2nd ed. PWS-Kent: Boston.
- Hardle, W. (1997). *Applied Non Parametric Regression*. Cambridge University Press: Cambridge.
- Hooge R D, Paul De Deyn P. (2001) *Applications of the Morris water maze in the study of learning and memory*. *Brain Research Reviews* 36:60-90.
- Morris R. (1984). *Developments of water-maze procedure for studying spatial learning in the rat*. *J.Neuroscience Methods*.11:47-60.
- Robertson, T.W., Wright, F.T., and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. Chichister: Wiley.
- Van Dam D, Lenders G, Paul De Deyn P. (2006) *Effect of Morris water maze diameter on visual-spatial learning in different mouse strains* *Neurobiology of learning and Memory* 85: 164-177.

Memoria del 2º Encuentro Iberoamericano de
Biometría y la V Reunión de la Región
Centroamericana y del Caribe de la
Sociedad Internacional
de Biometría

ISBN 978-607-00-2984-4

Editado en Xalapa, Veracruz, México