


Análisis de contenido en SPSS y KALPHA: Procedimiento para un Análisis Cuantitativo Fiable con la Kappa de Cohen y el Alpha de Krippendorff

Manuel Goyanes

Universidad Carlos III de Madrid (España) ✉ 

Valeriano Piñeiro-Naval

Universidad de Salamanca (España) ✉ 

<https://dx.doi.org/10.5209/esmp.92732>

Recibido: 27 de noviembre de 2023 / Aceptado: 19 de enero de 2024

ES Resumen: Durante las últimas décadas, el análisis de contenido se ha revelado como una de las técnicas de investigación fundamentales en los estudios de comunicación y periodismo. Una amplia literatura ha descrito, con gran detalle, los procedimientos para su sistemática y correcta implementación, lo que ha desembocado en investigaciones empíricamente más robustas y precisas. Este avance en la aplicación de los protocolos metodológicos no se ha trasladado, de manera tan paulatina, a los cómputos y test estadísticos para establecer la fiabilidad entre codificadores en las publicaciones en español, a pesar de su central importancia a la hora de ofrecer evidencias científicas reproducibles. Esto se ha debido, entre otras razones, a la necesidad de calcular los indicadores en programas estadísticos, lo que ha obstaculizado su adopción generalizada. En este trabajo ilustramos, mediante SPSS y la macro KALPHA, la computación de dos pruebas estadísticas para medir el acuerdo entre codificadores: la Kappa (κ) de Cohen y el Alpha (α) de Krippendorff. El objetivo es, por ende, ofrecer y detallar a la comunidad académica hispánica un protocolo para generalizar la utilización de cómputos que permitan reportar, a investigadores sin un avanzado conocimiento estadístico, resultados fiables y estándares en el campo.

Palabras clave: análisis de contenido; acuerdo entre codificadores; Kappa de Cohen; Alpha de Krippendorff; SPSS; macro KALPHA.

ENG Content Analysis in SPSS and KALPHA: Procedure for a Reliable Quantitative Analysis with Cohen's Kappa and Krippendorff's Alpha

Abstract: During the last decades, content analysis has emerged as one of the main research techniques in communication and journalism studies. An extensive literature has described, in great detail, the procedures for its systematic and correct implementation, which has led to more empirically robust and precise research. This advance in the application of methodological protocols has not been transferred, in such a gradual manner, to the computations and statistical tests to establish reliability between coders in publications in Spanish, despite its central importance when it comes to offering reproducible scientific evidence. This has been due, among other reasons, to the need to calculate the indicators in statistical programs, which has hindered their widespread adoption. In this work we illustrate, using SPSS and the KALPHA macro, the computation of two statistical tests to measure the agreement between coders: Cohen's Kappa (κ) and Krippendorff's Alpha (α). The objective is, therefore, to offer and detail to the Hispanic academic community a protocol to generalize the use of computations that allow researchers without advanced statistical knowledge to report reliable and standard results in the field.

Keywords: content analysis; agreement between coders; Cohen's Kappa; Krippendorff's Alpha; SPSS; KALPHA macro.

Sumario: 1. Introducción. 1.1. κ de Cohen vs. α de Krippendorff. 2. La Kappa de Cohen (κ). 2.1. Suposiciones de la Kappa de Cohen. 2.2. Configuración de los datos. 2.3. Procedimiento con interfaz de usuario y con sintaxis SPSS. 2.4. Procedimiento: ruta con sintaxis SPSS. 2.5. Resultados. 2.6. Reporte estadístico estandarizado. 3. El Alpha de Krippendorff (α). 3.1. Supuestos del Alpha de Krippendorff. 3.2. Configuración de los datos. 3.3. Procedimiento de cálculo con interfaz de usuario y sintaxis SPSS. 3.4. Resultados. 3.5. Reporte estadístico estandarizado. 4. Consideraciones finales. 5. Notas. 6. Referencias.

Cómo citar: Goyanes, M., & Piñeiro-Naval, V. (2024). Análisis de contenido en SPSS y KALPHA: Procedimiento para un Análisis Cuantitativo Fiable con la Kappa de Cohen y el Alpha de Krippendorff. *Estudios sobre el Mensaje Periodístico* 30 (1), 123-140. <https://dx.doi.org/10.5209/esmp.92732>

1. Introducción

En un campo teórica y metodológicamente tan diverso como el de las ciencias de la comunicación, pocas técnicas de investigación y recolección de datos han sido tan relevantes como el análisis de contenido (Riffe & Freitag, 1997; Stemler, 2015; Walter *et al.*, 2018). Este patrón de dominio del mensaje como objeto de estudio y del análisis de contenido como técnica de investigación es, quizá, una de las características idiosincráticas que mejor definen la producción científica en español, como se puede observar en una miríada de estudios que reflexionan sobre la situación y la evaluación metodológica del campo (Caffarel-Serra *et al.*, 2017; Goyanes *et al.*, 2018; Piñeiro-Naval & Morais, 2019; Segado-Boj *et al.*, 2023). En este contexto, bien sea por su “facilidad” de ejecución debido a décadas de perfeccionamiento y transmisión de sus procedimientos, su carácter accesible gracias al que cualquier investigador puede ejecutarlo sin demasiados recursos materiales, o por su utilidad al describir, de forma sistemática, objetiva y cuantitativa, los mensajes divulgados en los medios de comunicación (Bardin, 1986; Piñuel, 2002), el análisis de contenido simboliza, mejor que ninguna otra técnica de investigación, la naturaleza, los orígenes y la evolución histórica de la investigación en comunicación.

A pesar de los avances y refinamientos teóricos y procedimentales para aplicar correctamente el análisis de contenido, el factor humano todavía sigue siendo clave para generar buenos resultados (Singletary, 1993; Lombard *et al.*, 2002). En general, esta técnica se basa en datos generados por individuos (de aquí en adelante, codificadores), con arreglo a preguntas de investigación que motivan algún tipo de juicio o toma de decisión (Neuendorf, 2002). Estos codificadores juzgan o interpretan el contenido, generalmente textual (aunque también audio-visual), procedente de periódicos, formatos televisivos y radiofónicos, sitios web, redes sociales o, incluso, artículos académicos, con el cometido general de realizar inferencias sobre el tipo de contenido, frecuencia o magnitud de un hecho (Hayes & Krippendorff, 2007; Neuendorf & Kumar, 2016).

Más allá del aspecto procedimental que, por otra parte, ya se ha pautado con gran detalle en manuales canónicos sobre el tema (véanse: Berelson, 1952; Berger, 2016; Igartua, 2006; Kolbe & Burnett, 1991; Krippendorff, 2004a; Riffe *et al.*, 2019; Wimmer & Dominick, 2011), uno de los factores clave que determina la calidad de un análisis de contenido (y, por lo tanto, su reproductibilidad) es el entrenamiento y la fiabilidad entre codificadores (Bos & Tarnai, 1999; Kolbe & Barnett, 1991; Lacy *et al.*, 2015; Lombard *et al.*, 2002; Lovejoy *et al.*, 2014; Singletary, 1993), al ser el análisis de contenido una empresa fuertemente dependiente del juicio humano (Potter & Levine-Donnerstein, 1999). En este sentido, la codificación “manual” de unidades de análisis es el procedimiento clave y, sin embargo, la conceptualización de las

categorías de análisis y el propio contenido manifiesto o latente examinado dista mucho de su fácil ejecución. Por ejemplo, la distancia entre la observación y la correcta inferencia depende directamente de la naturaleza del mensaje observado, siendo un análisis de contenido más fiable, *a priori*, cuanto más manifiesta (por ejemplo, el conteo de palabras) y menos latente (por caso, juicios sobre un significado) sea su interpretación.

En consecuencia, el interrogante que un análisis de contenido debe responder es: ¿los resultados obtenidos procedentes de juicios subjetivos humanos atienden a decisiones arbitrarias y no reproducibles por otros humanos, o corresponden a las características y propiedades de las unidades de análisis que otros individuos también podrían alcanzar? Esta pregunta hace referencia a la fiabilidad entre codificadores y a la necesidad de reportar pruebas estadísticas que garanticen, con un cierto margen de error, que los resultados alcanzados por los juicios de los codificadores puedan ser reproducibles. Estas pruebas, ciertamente necesarias y demandadas en las revistas científicas de nuestro campo, constituyen una evidencia cuantitativa con la que testar la fiabilidad de los resultados y, bien presentadas, encarnan la estructura que sostiene cualquier otra inferencia estadística procedente de esos datos. Ahora bien, ¿existen evidencias meta-investigativas acerca del uso y la generalización de dichas pruebas?

En el contexto internacional, Lovejoy *et al.* (2016) se ocuparon de observar la producción divulgada en *Journalism & Mass Communication Quarterly*, *Journal of Communication* y *Communication Monographs* desde 1985 hasta 2014. Para ello seleccionaron una muestra de 672 trabajos que empleaban el análisis de contenido, determinando que el reporte de la fiabilidad pasa del ~20% de los casos, en 1985, hasta el ~90%, en 2014. Este significativo aumento atestigua la relevancia de la fiabilidad entre codificadores, generalizada mediante el cómputo de parámetros como el Alpha (α) de Krippendorff y la Kappa (κ) de Cohen que son, en el período 2010-2014, los más comunes (28,8% y 25,2%, respectivamente).

Por el contrario, la situación en el ámbito hispánico es diferente. A este respecto, Piñeiro-Naval (2020) identificó, entre 2013 y 2017, un total de 262 artículos publicados en destacadas revistas españolas (*Comunicar*, *Profesional de la Información*, *Revista Latina de Comunicación Social* y *Communication & Society*) e hispanoamericanas (*Cuadernos.info*, *Palabra Clave* y *Comunicación y Sociedad*) donde se había empleado esta técnica. Sus hallazgos señalaban que el 69,8% de los estudios prescindían de cualquier estadístico que reflejase el acuerdo entre codificadores. Asimismo, se observaba que las cabeceras hispanoamericanas tendían a publicar estudios con reporte de fiabilidad en mayor proporción que las españolas, una circunstancia que refuerza la intención del presente trabajo: llenar un sensible vacío detectado en el campo en español.

1.1. κ de Cohen vs. α de Krippendorff

Desde un punto de vista estrictamente estadístico, el Alpha de Krippendorff suele considerarse la medida idónea para evaluar el acuerdo entre codificadores en un análisis de contenido porque reúne los mejores criterios y/o propiedades (Hayes & Krippendorff, 2007; Krippendorff, 2004b, 2011), con lo que se recomienda su utilización frente a la Kappa de Cohen (Cohen, 1960). Por decirlo de otro modo, el test de Krippendorff es ejecutable en todos los análisis de contenido estándares (análisis de contenido con más de dos codificadores, permutables o intercambiables, y con casos vacíos, por ejemplo), mientras que la Kappa de Cohen no. Es por lo que se explican detalladamente, en la siguiente sección, las suposiciones básicas de la Kappa de Cohen con el fin de que los futuros usuarios de esta prueba comprueben, antes de aplicarla, si cumplen o violan sus suposiciones. Pero si el α de Krippendorff suele valorarse con mejores propiedades, ¿por qué se sigue utilizando la κ de Cohen?

Dos son las razones fundamentales. En primer lugar, y a diferencia de lo que sucede con el α de Krippendorff, la κ de Cohen está incluida en la mayoría de los paquetes estadísticos más populares en ciencias sociales, como SPSS, lo que facilita su utilización por parte de la comunidad. Ello se debe a que la κ de Cohen es una contribución estadística simple pero bien diseñada, que mejoraba significativamente a sus antecesoras: la S de Bennett (Bennett *et al.*, 1954) y la Pi (π) de Scott (Scott, 1955). Asimismo, es relevante destacar que, desde su publicación en 1960, la κ de Cohen se ha ido utilizando sistemáticamente en relevantes estudios científicos de diferentes campos, lo que ha acrecentado su impacto y popularidad.

En segundo lugar, la κ de Cohen es una prueba simple, suficiente y muy útil que encaja en la mayoría de análisis de contenido estándar, es decir, aquellos en los que dos codificadores únicos evalúan todo el corpus de variables categóricas. Por el contrario, el α de Krippendorff es, en su origen, un test estadístico diseñado *ad hoc* para el análisis de contenido en comunicación, ciertamente amplio y más robusto, pero que ha requerido, hasta la fecha, de paquetes estadísticos diseñados *ex profeso* para su computación, como por ejemplo el "ReCal" (Freelon, 2013) o la macro "KALPHA" para SPSS (Hayes & Krippendorff, 2007), entre otros. En consecuencia, esta falta de inclusión en los principales paquetes estadísticos ha dificultado significativamente su uso y adopción frente a otras pruebas con quizá peores características.

2. La Kappa de Cohen (κ)

Como ya hemos sugerido, la Kappa de Cohen es una medida de acuerdo entre codificadores para variables categóricas cuando hay dos codificadores involucrados. Un ejemplo podría ser el siguiente: estamos interesados en saber, dentro de una muestra de estudios de una revista, cuántos de esos artículos están firmados, como primer autor, por investigadores españoles. En este caso, la variable objeto de estudio es "nacionalidad del primer firmante", que tendría dos niveles: "español" o "no español". El acuerdo entre los dos codificadores encargados de hacer el análisis de contenido se realizaría con la Kappa de

Cohen y nos daría un coeficiente de fiabilidad, teniendo en cuenta la proporción de acuerdo aleatorio (alcanzado por azar):

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

donde $Pr(a)$ sería la proporción de acuerdo observado, a saber, la proporción de casos en el que los dos codificadores están de acuerdo; mientras que $Pr(e)$ supondría la proporción hipotética de acuerdo aleatorio, esto es, la proporción de casos donde se esperaría que los dos codificadores estuvieran de acuerdo simplemente por casualidad. El coeficiente Kappa de Cohen tiene un rango de -1 a 1, siendo los valores más altos un indicativo de mejor nivel de fiabilidad. En consecuencia, el testeo de la hipótesis sería:

H_0 : $\kappa = 0$, el acuerdo entre codificadores no es diferente del acuerdo aleatorio;

H_A : $\kappa \neq 0$, el acuerdo entre codificadores es diferente del acuerdo aleatorio.

2.1. Suposiciones de la Kappa de Cohen

Para ejecutar este test estadístico es necesario cumplir las siguientes cinco suposiciones. En caso de que se viole una o varias de estas premisas, el test no debería ejecutarse. Sin embargo, como observaremos, estas suposiciones son ampliamente cumplidas por la mayoría de análisis de contenido estándar del campo.

1. La variable o variables analizadas deben ser medidas como categóricas (ordinales o nominales) y mutuamente excluyentes. En el ejemplo anterior, el primer firmante del artículo es "español" o "no español", lo que significa que los valores no se superponen: un juicio no puede ser "español" y "no español" al mismo tiempo. Por ende, solo un valor puede ser seleccionado para cada codificación.
2. Los juicios sobre las preguntas de investigación que hacen los dos codificadores son los mismos, es decir, las observaciones son pareadas. En otras palabras, los codificadores realizan juicios subjetivos de las mismas observaciones. En el ejemplo anterior, si hay una muestra de 50 artículos, significaría que el Codificador A y el Codificador B realizan la codificación de la nacionalidad de los 50 artículos. Debido al tiempo necesario para implementar observaciones pareadas en análisis con grandes muestras, la investigación reciente ha desarrollado estrategias para simplificar este requerimiento. La más extendida es que el Codificador A realice la codificación total de la muestra y el Codificador B, por su parte, se encargue de un porcentaje dado de observaciones aleatorias de la muestra para luego contrastar. En el caso anterior, el Codificador A se ocuparía de toda la muestra de artículos ($N = 50$), mientras que el codificador B observaría el 15-20% de la muestra ($N = 7.5$ o 10); o sea, entre 8 y 10 estudios seleccionados de modo aleatorio. A pesar de que este procedimiento viola una de las suposiciones básicas de la prueba, su ejecución es común o estándar en el campo. En cualquier

caso, se recomienda realizar codificaciones con observaciones pareadas, especialmente cuando la muestra y las variables objeto de estudio son pocas o con contenido manifiesto.

3. Cada variable debe contar con los mismos valores de respuesta y la tabulación cruzada (tabla de contingencia) debe ser simétrica. Dicho de otro modo, los valores de cada variable tienen que ser los mismos para cada codificador. En el caso anterior los valores son dos: “español” y “no español”, que se representan en una tabla de contingencia 2x2. Si hubiese tres valores para la variable (“español”, “francés” y “portugués”), y dos codificadores, la tabla de contingencia sería 3x3; esto es, los mismos tres valores simétricos para los dos codificadores involucrados.
4. Los dos codificadores encargados de hacer juicios son independientes, de manera que el Codificador A realiza sus observaciones al margen de las que efectúa (o haya efectuado) el Codificador B. Para evitar cualquier tipo de sesgo en este sentido, si los codificadores realizan sus juicios en un mismo lugar, sería conveniente que lo hiciesen en espacios, aulas o habitaciones separadas y que no discutiesen sus apreciaciones. Asimismo, se recomienda que cada una de las codificaciones, si se hacen en un mismo espacio, se hagan en horarios diferentes para que los codificadores no coincidan, salvo que el objeto de la codificación sea algún contenido web susceptible de volatilizarse y que no se pueda descargar.
5. El número de codificadores que hacen juicios de todas las observaciones asciende a dos. En

consecuencia, si para el análisis se utilizan más de dos codificadores o se han utilizado codificadores diferentes para cada observación, la Kappa de Cohen no es el estadístico apropiado y otras pruebas como el Alpha de Krippendorff o la Kappa de Fleiss (Fleiss, 1971) deben ser utilizadas.

Una vez demostrado el cumplimiento de las suposiciones, el siguiente paso corresponde a la configuración de los datos para ejecutar el estadístico. A continuación, describimos este proceso para el paquete estadístico SPSS.

2.2. Configuración de los datos

El supuesto utilizado para esta guía práctica es la codificación de 50 artículos por parte de dos codificadores.¹ En concreto, estos deben realizar un juicio, atendiendo a la filiación de los primeros firmantes de los 50 artículos, y determinar si son “español” o “no español”; por tanto, españoles o extranjeros. Para introducir variables y valores en SPSS, se recomienda la lectura de una guía de iniciación para familiarizarse con el *software*, ya que es necesario para configurar la matriz de datos.

Una vez codificadas las observaciones, esta base de datos consistirá en dos variables (una variable para cada codificador: Codificador A y Codificador B), con dos niveles o valores (“español” o “no español”) y 50 observaciones (correspondientes a los juicios que los codificadores deben hacer para los 50 artículos que conforman la muestra). Así pues, la matriz finalizada debería asemejarse a lo reflejado en la Imagen 1.²

Imagen 1. Vista previa de 13 de los 50 juicios de los 2 codificadores que han efectuado el análisis de contenido (variable: “procedencia del primer firmante”)

	Codificador_A	Codificador_B	var	var	var	var	var	var	var	var	var	var
1	Español	Español										
2	Español	Español										
3	Español	Español										
4	Español	Español										
5	Español	Español										
6	Español	Español										
7	No español	No español										
8	No español	Español										
9	Español	Español										
10	Español	Español										
11	Español	Español										
12	Español	Español										
13	Español	Español										

2.3. Procedimiento con interfaz de usuario y con sintaxis SPSS

Al igual que la mayoría de los paquetes estadísticos más utilizados en ciencias sociales, SPSS cuenta con una interfaz de usuario interactiva para facilitar la ejecución de los análisis estadísticos, complementada con un archivo de sintaxis en el que el usuario puede escribir código o comandos para ejecutar

directamente los análisis. Tanto la ruta interactiva con interfaz de usuario como la escritura de sintaxis son rutas disponibles para la Kappa de Cohen y su ejecución en SPSS. Las dos rutas tienen ventajas y desventajas, que pasamos a describir en las siguientes secciones.

El procedimiento para ejecutar la Kappa de Cohen es similar a realizar un análisis de contingencia

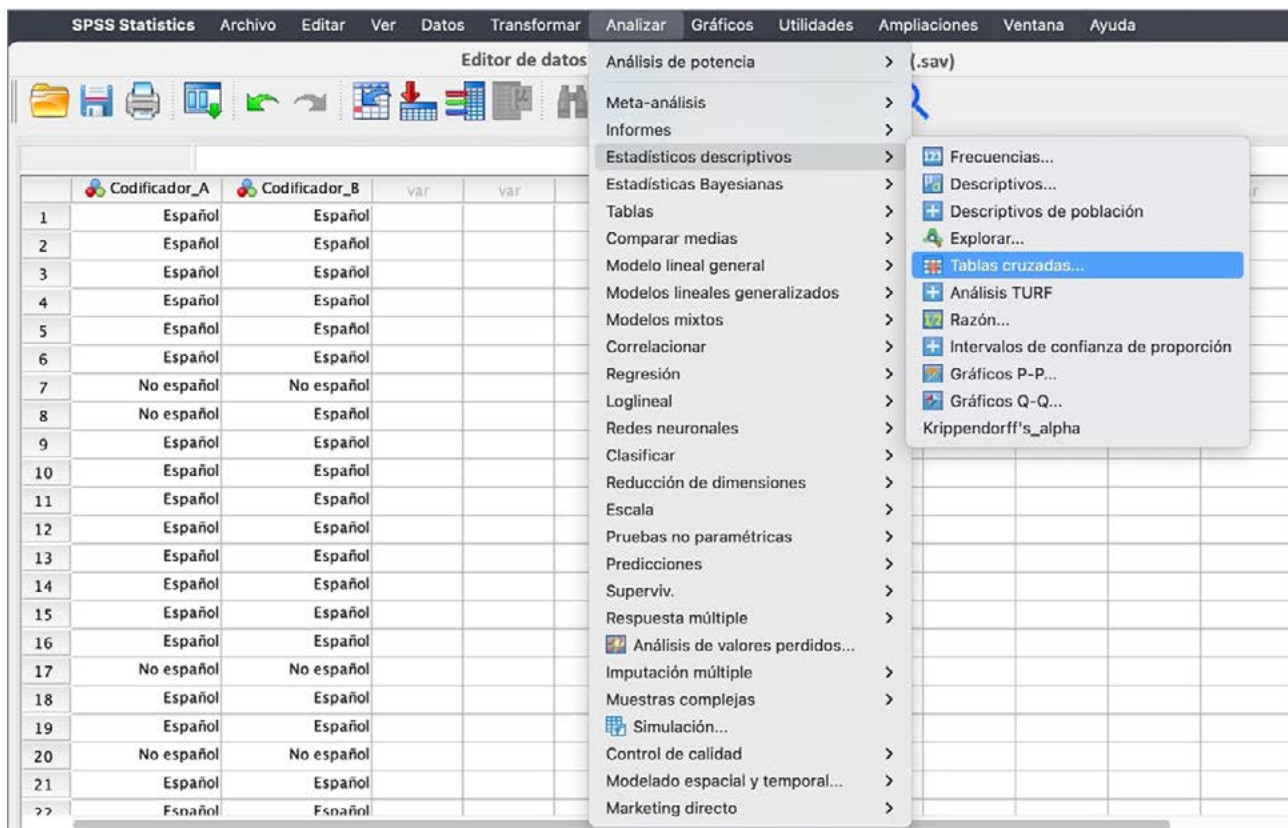
clásico (*crosstabs*, para el *software* en inglés). La ruta u opción más sencilla es a través de la interfaz del programa, ya que permite al usuario establecer una comunicación más sencilla, y sin código, con el paquete estadístico. Por el contrario y como desventaja, cada vez que el usuario quiera ejecutar el estadístico y haya hecho algún cambio en las variables (su orden, por ejemplo), el usuario deberá volver a realizar las operaciones que se describen a continuación. Es por ello por lo que, siempre que se pueda, recomendamos utilizar la sintaxis, guardarla y tenerla como *back-up*, por si en el futuro hubiera que volver a ejecutar o revisar los análisis.

En el caso de la ejecución de la Kappa de Cohen son muy pocas las variables que se consideran (únicamente dos: Codificador A y Codificador B) y, por lo tanto, la probabilidad de realizar cambios en ellas es limitada, pero en caso de análisis más complejos,

como la regresión jerárquica, es prescriptivo (y casi obligatorio) guardar la sintaxis para dejar constancia del orden y las variables introducidas y poder, así, replicar y revisar los análisis en el futuro. En cualquier caso, bien se utilice la interfaz de usuario o la sintaxis directamente, resulta útil guardar en un documento la sintaxis de los análisis ejecutados con el fin de saber qué y cómo se han hecho y, en un momento dado, que se puedan corregir si se observa algún un error.

Como hemos mencionado anteriormente, a diferencia de lo que ocurre con el Alpha de Krippendorff, que no viene incluida en la mayoría de los paquetes estadísticos, la Kappa de Cohen sí. En concreto, el procedimiento o ruta en SPSS para computar este estadístico desde la interfaz de usuario es: “Analizar / Estadísticos descriptivos / Tablas cruzadas”, como se refleja en la Imagen 2:

Imagen 2. Ruta para ejecutar la Kappa de Cohen en SPSS



Una vez desplegado el cuadro de diálogo de la tabla de contingencia, hay que pasar la variable “Codificador A” a la “Fila(s)” y “Codificador B” a la “Columna(s)”, o viceversa, y así crear una tabla de contingencia 2x2 clásica (véase Imagen 3). Al ser una tabla de contingencia 2x2 el orden de los factores no afecta a los resultados y, por ende, se pueden intercambiar en filas o columnas las

variables anteriores. Una vez introducidas hay que seleccionar el botón “Estadísticos” y, del panel de opciones, “Kappa” (véase Imagen 4) para clicar luego en “Continuar”. Por último, hemos de marcar el botón “Celdas” y dejar activados los cuadros “Observadas”, “Esperadas”, “Fila”, “Columna” y “Total” (véase Imagen 5), para finalmente hacer clic en “Pegar” (véase Imagen 6).

Imagen 3. Vista previa de la tabla de contingencia

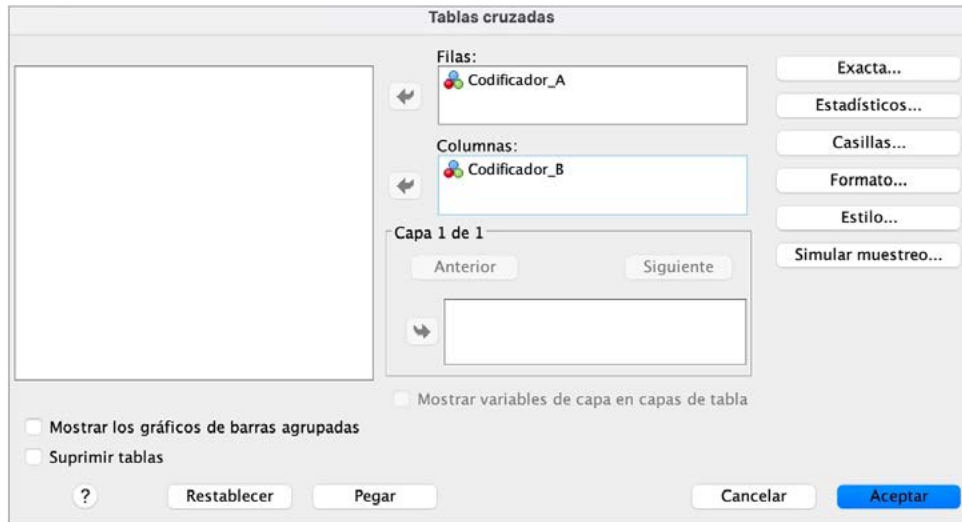


Imagen 4. Vista previa de la selección de la Kappa de Cohen

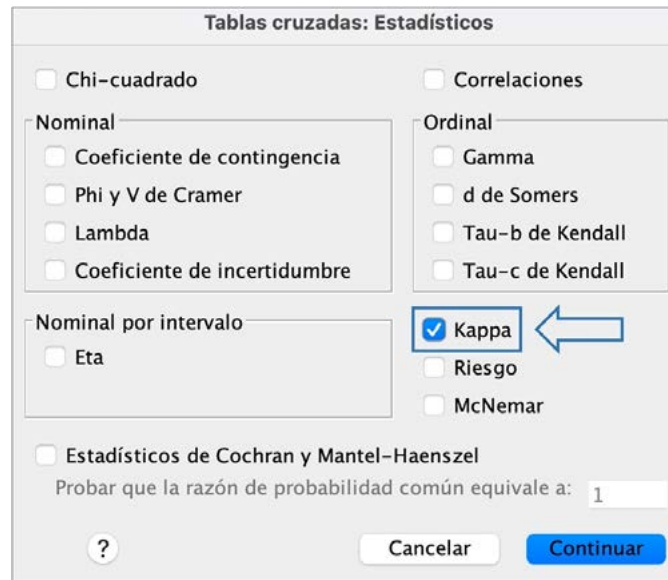


Imagen 5. Vista previa de las características del análisis que se deben dejar marcadas

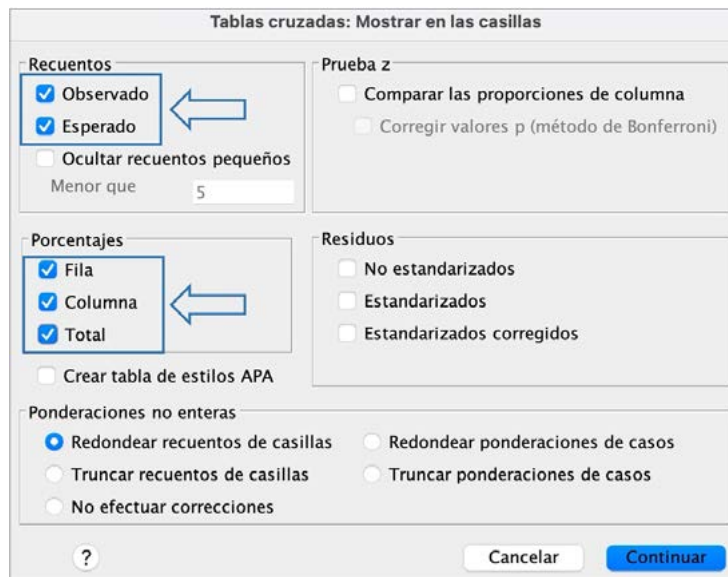
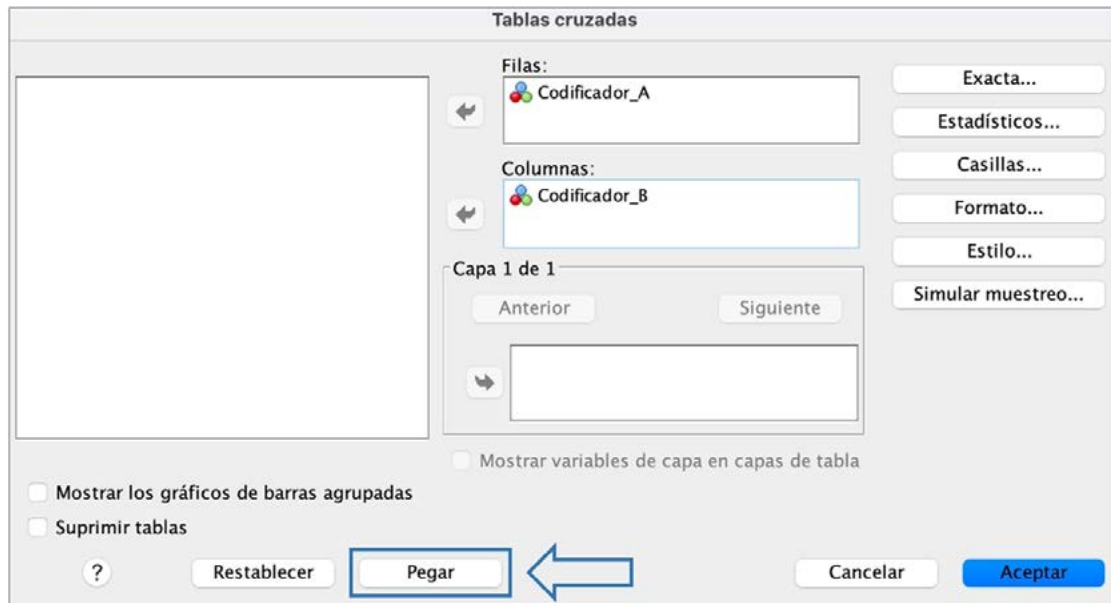


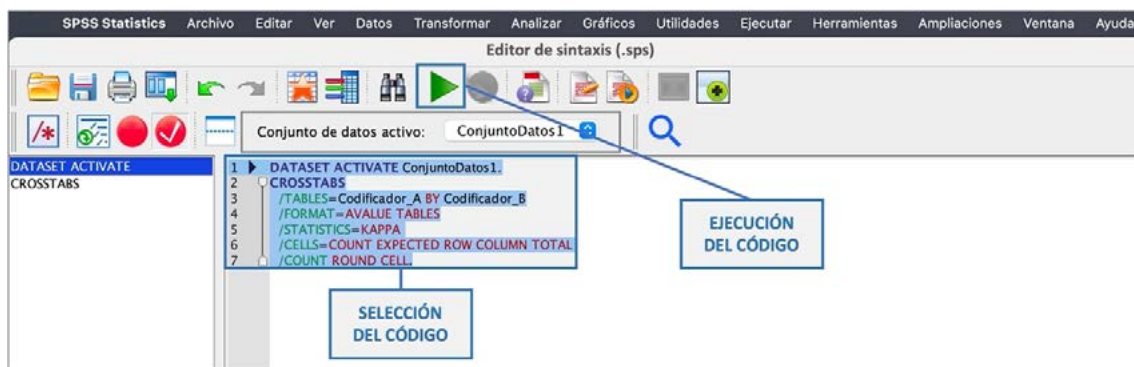
Imagen 6. Vista previa del comando "Pegar"



La selección "Pegar" ejecuta la sintaxis de la ruta y las opciones seleccionadas en la descripción anterior. Por lo tanto, al "Pegar" lo que hemos generado es una sintaxis o código procedente de la selección del estadístico y las variantes que hemos marcado. De este modo, "Pegar" no significa la ejecución de la prueba estadística, sino la generación de la sintaxis para poder ejecutarla ahora y en el futuro. El usuario, para generar la prueba, ha de seleccionar con el ratón esa sintaxis y pinchar acto seguido en "ejecutar selección", en el icono triangular de color verde que aparece en el panel

de herramientas del documento de la sintaxis (véase Imagen 7). De esta forma, el usuario se asegura tener un registro de los análisis ejecutados, con sus procesos y opciones. Por tanto, los análisis podrán ser repetidos más adelante (quizá 5 meses después, cuando se haya recibido la primera revisión del artículo) y mantener la naturaleza de los análisis. Si el usuario quiere obviar la sintaxis, en lugar de clicar en "Pegar" puede hacerlo en "Aceptar" y la prueba estadística se habrá ejecutado directamente en el *output* de SPSS, pero sin haber generado ni guardado la sintaxis.

Imagen 7. Vista previa de la selección y ejecución de la sintaxis de la Kappa de Cohen.



2.4. Procedimiento: ruta con sintaxis SPSS

Al igual que en cualquier otro paquete estadístico, el SPSS tiene sintaxis y comandos que facilitan la ejecución de las pruebas que se quieran realizar. Esta sintaxis se emplea cuando se conoce tanto el código como los test que SPSS genera, y suele ser aplicada directamente si el usuario tiene un conocimiento avanzado del paquete. Por el contrario, para los investigadores que están iniciando su empleo de SPSS se recomienda la ruta anterior y "Pegar" la sintaxis para generarla y luego ejecutarla.

Los usuarios más avanzados que quieran escribir la sintaxis y ahorrarse la ruta a través de la interfaz de usuario deberán reproducir el siguiente código:

CROSSTABS

```
/TABLES=Codificador_A BY Codificador_B
/FORMAT=AVALUE TABLES
/STATISTICS=KAPPA
/CELLS=COUNT EXPECTED ROW COLUMN
TOTAL
/COUNT ROUND CELL.
```

Como observamos, la sintaxis corresponde a un análisis con tablas de contingencia clásico al que se le añaden las opciones adicionales de "Kappa", "Count", "Expected", "Row", "Column" y "Total", correspondientes al *output* que nos gustaría que el *software* nos ofreciese y que se explicará

a continuación. La primera fila de la sintaxis ejecuta la tabla de contingencia, en este caso, 2x2, con “Codificador_1 BY Codificador_2”. En este ejemplo el nombre de las variables es “Codificador_1” y “Codificador_2”, que deben ser adaptadas a los nombres que cada usuario haya usado en su configuración de datos.

En las siguientes secciones interpretamos los resultados y ofrecemos una guía de cómo reportarlos de manera estandarizada.

2.5. Resultados

La primera tabla, el “Resumen de Procesamiento de Casos” (Imagen 8), aporta un resumen descriptivo de los juicios de los codificadores. En particular, la tabla da un resumen de: casos “Válidos”, esto es, los que se han considerado para la ejecución de la Kappa de Cohen; casos “Perdidos”, es decir, las observaciones que no se han codificado; y casos “Totales”, a saber, la suma de casos perdidos y válidos.

Imagen 8. Sumario de procesamiento de casos

Tablas cruzadas

Resumen de procesamiento de casos

	Válido		Casos Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Codificador_A * Codificador_B	50	100,0%	0	0,0%	50	100,0%

La siguiente imagen (Imagen 9), llamada tabla de contingencia o cruzada, ofrece el cruce de juicios para los dos codificadores teniendo en cuenta los dos valores “español” y “no español”. La tabla, de acuerdo con el nombre de nuestras dos variables, aporta los resultados de los juicios de “Codificador_A” y “Codificador_B”, pero estos nombres serán diferentes si el usuario, durante el proceso de configuración de sus datos, ha utilizado otros.

La tabla de contingencia (Imagen 9) muestra resultados que nos permiten entender el nivel de acuerdo o desacuerdo entre codificadores al efectuar sus juicios. Específicamente, el nivel de acuerdo (o número de veces pareadas donde los dos codificadores han estado de acuerdo en sus juicios) se puede examinar observando la diagonal que va de izquierda a derecha y de arriba abajo en la Imagen 9, destacada en amarillo. La tabla sugiere que, de los

50 juicios efectuados, en 48 de ellos los dos codificadores han estado de acuerdo (43 + 5), correspondientes a 43 juicios en que han estado de acuerdo en codificar a los autores como “español” y 5 en que han coincidido en codificar a los autores como “no español”.

Si observamos la diagonal contraria (la que va de derecha a izquierda y de arriba abajo, destacada en turquesa en el Imagen 9), veremos el número de veces en el que los codificadores no se han puesto de acuerdo (1 + 1), correspondientes a un juicio en el que el Codificador_A ha decidido que la observación (la procedencia del autor) era “no español” y que, sin embargo, el Codificador_B ha considerado “español”; y otra observación en la que el Codificador_A ha decidido que la observación era “no español”, mientras que el Codificador_B ha considerado “español”.

Imagen 9. Tabla cruzada 2x2 que explica el acuerdo entre codificadores (valores del ejemplo).

Tabla cruzada Codificador_A*Codificador_B

Recuento		Codificador_B		Total
		Español	No español	
Codificador_A	Español	43	1	44
	No español	1	5	6
Total		44	6	50

Nota: hemos eliminado los valores esperados y los porcentajes de columna, fila y total con el objetivo de presentar los datos de manera simplificada.

En un sentido general, este tipo de tabla contendría la siguiente información:

Tabla 1. Tabla de contingencia 2x2 que explica el acuerdo entre codificadores (información general)

		Observador B		Total
		Positivo	Negativo	
Observador A	Positivo	a = positivos verdaderos	b = falsos negativos	r
	Negativo	c = falsos positivos	d = negativos verdaderos	s
Total		t	u	N

Teniendo en cuenta los datos anteriores, es posible computar fácilmente un porcentaje de acuerdo entre codificadores (Fleiss *et al.*, 2003) mediante la suma de todos los acuerdos (diagonal amarilla: 43 + 5 = 48), y la división por el total del número de juicios efectuados (N = 50). Esto se traduce, a nivel matemático, en la Proporción de acuerdo Observado:

$$PO = \frac{a + d}{a + b + c + d} * 100.$$

En nuestro caso, PO = 48 ÷ 50 = 0.96 o 96% (0.96 x 100). Así pues, los codificadores han estado de acuerdo en el 96% de los juicios o, lo contrario, han estado en desacuerdo un 4% de las veces. Este porcentaje, ciertamente sesgado porque no toma en cuenta el acuerdo aleatorio (Artstein & Poesio, 2008), debe ser reportado junto al coeficiente Kappa de Cohen ofrecido a continuación (Imagen 10).

Imagen 10. Coeficiente de la Kappa de Cohen para el ejemplo empleado

Medidas simétricas					
		Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Medida de acuerdo	Kappa	.811	.130	5,732	<.001
N de casos válidos		50			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

En concreto, la Imagen 10 refleja el coeficiente de la Kappa de Cohen, que en el ejemplo expuesto a lo largo de este artículo es de “.811” (marcado en amarillo en la Imagen 10), correspondiente al acuerdo entre codificadores más allá del acuerdo casual que pudiera existir entre ellos. Como se ha explicado anteriormente, el rango de esta prueba estadística es de -1 a +1, siendo el valor -1 indicativo de falta total de acuerdo (los revisores no se han puesto de acuerdo en ninguna observación), 0, indicativo de que el acuerdo no ha sido más elevado que el acuerdo azaroso, y +1, indicativo del acuerdo total entre los dos codificadores. Conociendo estos datos, se asume lógicamente que cuanto más altos sean los valores del coeficiente mejor será la fiabilidad de la codificación, pero ¿qué rangos son los más aceptables?

La literatura ha propuesto rangos de la Kappa de Cohen, ya ampliamente aceptados, que explican la fortaleza de la fiabilidad entre codificadores. En concreto, los valores propuestos inicialmente por Altman (1999) corresponden a:

- <0.20 = Fiabilidad entre codificadores pobre
- 0.21 – 0.40 = Fiabilidad entre codificadores discreta
- 0.41 – 0.60 = Fiabilidad entre codificadores moderada
- 0.61 – 0.80 = Fiabilidad entre codificadores sustancial
- 0.81 – 1.00 = Fiabilidad entre codificadores muy buena

Más allá del reporte del coeficiente Kappa, es fundamental también testar la hipótesis y, en consecuencia, ofrecer el valor *p* que genera el estadístico. En concreto, de acuerdo con lo explicado anteriormente, el objeto de un buen análisis de contenido es H_A: κ ≠ 0; a saber, que el acuerdo entre codificadores sea diferente al acuerdo aleatorio. Por ende, cuando el valor *p* (marcado en verde en la Imagen 10) es menor a “.05” (*p* < .05), el coeficiente ofrece un resultado estadísticamente significativo (κ ≠ 0). En nuestro caso, como *p* < .001, se confirma la hipótesis alternativa y se puede afirmar que la fiabilidad entre codificadores es diferente de 0 y, además, viendo el valor del coeficiente, que la fiabilidad es muy buena.

De modo complementario, y para ofrecer unas evidencias más robustas de la fiabilidad entre codificadores, podemos reportar el Intervalo de Confianza (IC) al 95% del coeficiente Kappa. Para ello debemos efectuar unas computaciones, bastante simples, ya que el *output* del SPSS no lo proporciona. Primero tenemos que conocer el error estándar de la Kappa, que en nuestro caso es el “Error Estándar Asintótico” (marcado en color turquesa en la Imagen 10), correspondiente a “.130”. En muestras grandes se suele asumir que este error estándar se distribuye de manera normal y, en consecuencia, podemos computar el intervalo de confianza al 95% con la tabla de valores *z*, teniendo en cuenta que el error estándar representa aproximadamente el 68% del intervalo de confianza de la desviación estándar de la Kappa, lo

que significa que la tenemos que multiplicar por 1.96 para obtener el intervalo de confianza al 95%.

En nuestro caso, esta multiplicación ofrece el valor de “.254”, resultado de multiplicar .130 x 1.96. Por lo tanto, el intervalo de confianza al 95% de nuestro coeficiente es $.811 \pm .254$, lo que de modo estandarizado se reportaría como: .811 (95% IC, .557 a 1.065). Como vemos en este caso, el coeficiente Kappa, cuando se le suma el intervalo de confianza, ofrece un valor mayor que 1. Asimismo, en muchas ocasiones, el valor menor del intervalo de confianza estimado puede contener 0 y, no obstante, obtener un coeficiente estadísticamente significativo. Ello se debe a que el error estándar, para hacer los cálculos anteriores al 95%, no es el mismo que el error estándar para computar la significancia estadística (p) del coeficiente Kappa.

2.6. Reporte estadístico estandarizado

El reporte, en un sentido general, deberá realizarse en la sección metodológica del estudio, concretamente donde se describa el procedimiento para el análisis de contenido y el libro de códigos. A modo de ejemplo, mostramos los resultados siguiendo un modelo estandarizado del análisis anterior:

“Se ejecuta la Kappa de Cohen (Cohen, 1960), para determinar la fiabilidad entre los dos codificadores encargados de realizar el análisis de contenido. En concreto, dos codificadores han emitido juicios de una muestra de 50 artículos para determinar si el primer autor era «español» o «no español». Del total de 50 juicios, los codificadores han estado de acuerdo en 48 casos y en desacuerdo en 2, ofreciendo un porcentaje de acuerdo del 96%. Para determinar la fiabilidad entre los dos codificadores teniendo en cuenta la probabilidad de acuerdo aleatorio, se ejecuta la Kappa de

Cohen, ofreciendo un coeficiente de: $\kappa = .811$ (IC 95%: .557, 1.065), $p < .001$, sugiriendo una fiabilidad entre codificadores muy buena”.

3. El Alpha de Krippendorff (α)

Como hemos indicado anteriormente, el Alpha de Krippendorff se considera, desde el punto de vista estadístico, la mejor medida para calcular el acuerdo entre codificadores en un análisis de contenido (Hayes & Krippendorff, 2007; Krippendorff, 2004b, 2011). Entre otras razones, este parámetro puede emplearse con independencia del número de jueces involucrados en el proceso de codificación (dos o más), del nivel de medida de las variables incluidas en el libro de códigos (cualitativas o cuantitativas), del tamaño de la muestra (N) o de la posible existencia de valores perdidos (celdas vacías) en alguna de las observaciones.

Con respecto a esta última particularidad, lo más recomendable es que el investigador responsable del análisis de contenido verifique, antes del cruce de los datos de los codificadores, que no existen errores de grabación (por ejemplo, a causa del empleo de un código no contemplado en la categorización de una variable) o valores *missing* que, por descuidos involuntarios, se hayan filtrado en la matriz. Para ello, un procedimiento mucho más rápido que la verificación de las celdas una a una consistiría en solicitar a SPSS una tabla de frecuencias de cada variable y, en el supuesto de que exista algún valor extraño, este se podría detectar en el editor de datos (archivo .sav) mediante la selección de la columna donde aparezca la variable en cuestión y accediendo a: “Menú / Editar / Buscar y reemplazar” (Imagen 11). A partir de ahí podríamos dar con el valor que no procede y sustituirlo convenientemente (Imagen 12) con arreglo a una nueva codificación, que deberá efectuar el juez cuyo dato se haya perdido.

Imagen 11. Ruta para detectar errores de grabación (en este caso, del Codificador 1)

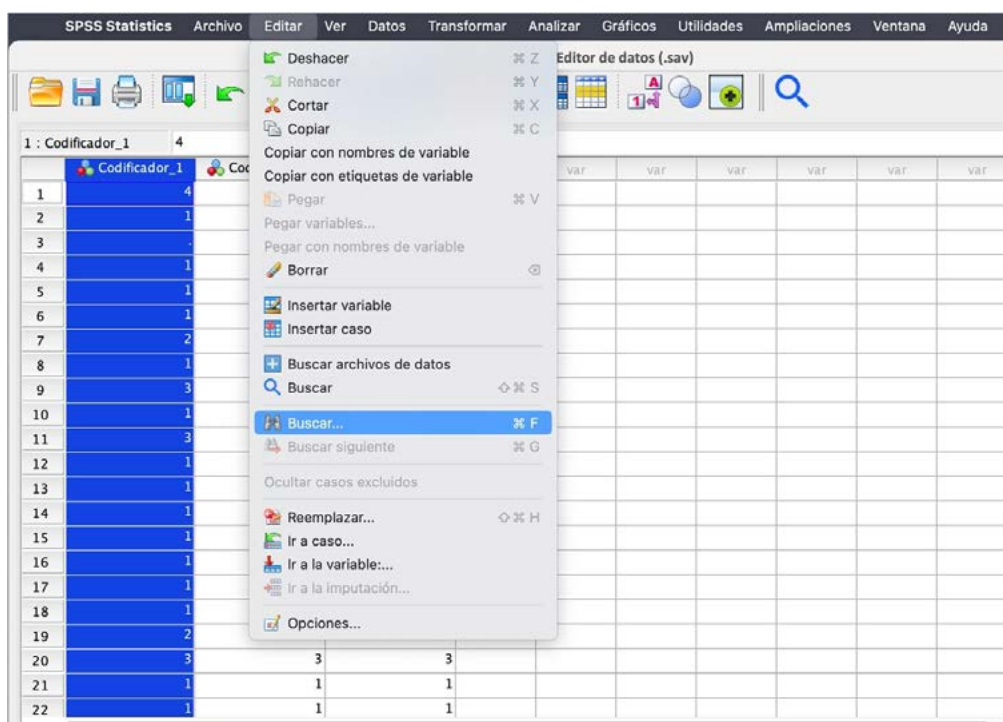
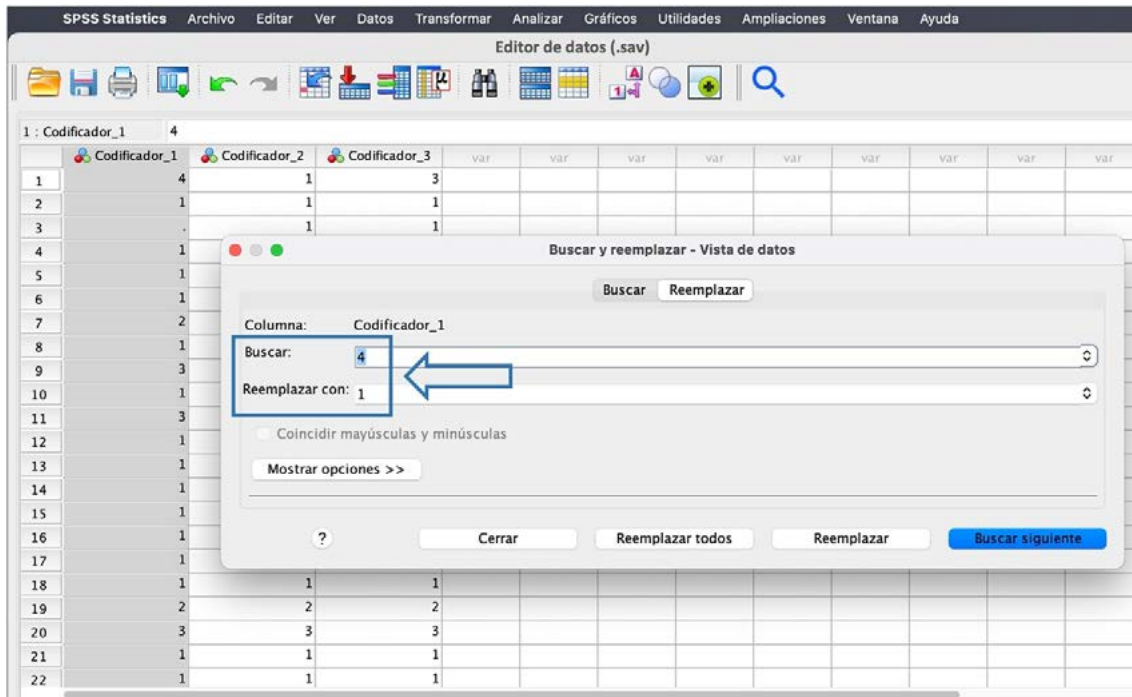


Imagen 12. Menú “Buscar y reemplazar” (el valor 4 no es una categoría real de la variable, por lo deberemos sustituirlo por 1)



En lo tocante a la formulación del Alpha de Krippendorff, esta es su versión simplificada (Manterola *et al.*, 2018):

$$\alpha = 1 - \frac{D_o}{D_e}$$

donde D_o es el desacuerdo observado y D_e el desacuerdo esperado por azar. El rango de este parámetro va de 0 a 1 y, según Krippendorff (2004a, 2004b), un resultado de “.667” sería el límite inferior admisible, e igual o superior a “.800” lo más adecuado. A partir de estas indicaciones, la comprobación de la hipótesis se efectuaría tal que así:

H_0 : $\alpha < .667$, el acuerdo entre codificadores no es diferente del acuerdo por azar;

H_A : $\alpha \geq .667$, el acuerdo entre codificadores es diferente del acuerdo por azar.

En el siguiente apartado resumiremos las condiciones necesarias para el empleo del Alpha de Krippendorff.

3.1. Supuestos del Alpha de Krippendorff

Para llevar a cabo esta prueba deben confluír 3 supuestos principales, como son:

1. La naturaleza de la(s) variable(s) en cuestión puede ser tanto cualitativa (nominal u ordinal) como cuantitativa (de intervalo o de razón). Si se cruzan variables cualitativas deben ser, al igual que sucede con κ , mutuamente excluyentes; si son de intervalo, deberán contemplar el mismo rango de valores (por caso, una escala tipo Likert de 1 a 5 puntos); y, si son de razón, serán cifras que podrán ir de 0 a $\pm \infty$ (por ejemplo, el número de citas que ha recibido un artículo o la cantidad de referencias bibliográficas que cita). A modo ilustrativo,

aquí emplearemos un nuevo ejemplo, consistente en una variable nominal multicategorica referente al tipo de artículo publicado; esto es: “empírico”, “teórico-ensayístico” y “metodológico” (se excluyen, por tanto, editoriales y reseñas del propio muestreo), de forma que los valores no se solaparían en la posterior codificación.

2. Los dos (o más) jueces implicados en el proceso de codificación efectúan inferencias subjetivas de las mismas observaciones. Siguiendo con el ejemplo anterior, si se trabaja con una muestra de 50 artículos y un equipo de 3 codificadores, tanto el Codificador 1, como el Codificador 2 y el Codificador 3 examinan los 50 artículos para determinar su tipología (empíricos, teórico-ensayísticos o metodológicos). Si la muestra fuera de un tamaño superior (pensemos en $N = 300$ artículos), se podría efectuar un reparto equitativo entre los 3 codificadores, asignándoles a cada uno de ellos una submuestra aleatoria de $N = 100$ manuscritos. A continuación, se volvería a seleccionar un 20% de cada submuestra ($N = 20$ artículos) para que los restantes codificadores también la analizaran, de modo que cada uno se ocuparía de 140 artículos ($100 + 20 + 20$; es decir, el Codificador 1 revisaría sus propios 100, a los que se sumarían 20 del Codificador 2 y otros 20 del Codificador 3, y así sucesivamente para los otros dos jueces). Este procedimiento permitiría obtener 2 matrices de datos: una general compuesta por 300 filas y una columna (el tipo de artículo), que se destinaría a efectos del reporte de resultados y que sería fruto de la fusión de las 3 submuestras (100 del Codificador 1 + 100 del Codificador 2 + 100 del Codificador 3); y otra compuesta por 60 filas (el 20% de la muestra) y 3 columnas (tipo de artículo según Codificador 1, Codificador 2 y Codificador 3), que permitirá efectuar el cruce de las observaciones de los 3

jueces para conocer su grado de acuerdo al asignar una categoría (empírico, teórico-ensayístico y metodológico) a los 60 artículos aleatoriamente seleccionados. Esta segunda base de datos se destinaría exclusivamente al cálculo del Alpha de Krippendorff.

- Los dos o más codificadores encargados de hacer juicios son independientes, intercambiables y libremente permutables. De nuevo, el trabajo de uno no puede interferir en el de los demás. A tal efecto, aconsejamos que el libro de códigos esté lo más claramente redactado posible y llevar a cabo un entrenamiento o pilotaje previo con los codificadores para aunar criterios y esclarecer eventuales dudas. En este sentido, será necesario que recopilemos un reducido número de casos equivalentes a las unidades de análisis, pero que no figuren en la muestra final para no sesgar los resultados. En nuestro ejemplo, bastaría con 5 artículos procedentes de revistas ajenas al

alcance del estudio o publicados en años no englobados en el lapso temporal examinado.

3.2. Configuración de los datos

Continuaremos con el supuesto utilizado a lo largo de esta guía práctica, consistente en la codificación de 50 artículos por parte, en esta ocasión, de 3 codificadores. Desde este planteamiento, los 3 han de inferir, a tenor de la naturaleza de los 50 manuscritos, si se trata de trabajos empíricos, teórico-ensayísticos o metodológicos. Así pues, la base de datos que crearemos a efectos del chequeo del Alpha de Krippendorff estará compuesta por 3 variables (una para cada codificador: Codificador 1, Codificador 2 y Codificador 3) con 3 niveles o categorías (“empíricos”, “teórico-ensayísticos” o “metodológicos”) y 50 filas (tantas como artículos figuran en la muestra). De este modo, la matriz de datos se asemejará a lo plasmado en la Imagen 13.

Imagen 13. Vista previa de 13 de las 50 observaciones a cargo de los 3 codificadores que han efectuado el análisis de contenido (variable “tipo de artículo”)

	Codificador_1	Codificador_2	Codificador_3	var	var	var	var	var	var	var
1	empírico	empírico	metodológico							
2	empírico	empírico	empírico							
3	.	empírico	empírico							
4	empírico	teórico-ensayístico	empírico							
5	empírico	empírico	empírico							
6	empírico	empírico	empírico							
7	teórico-ensayístico	teórico-ensayístico	teórico-ensayístico							
8	empírico	.	empírico							
9	metodológico	empírico	empírico							
10	empírico	empírico	empírico							
11	metodológico	metodológico	metodológico							
12	empírico	empírico	empírico							
13	empírico	empírico	empírico							

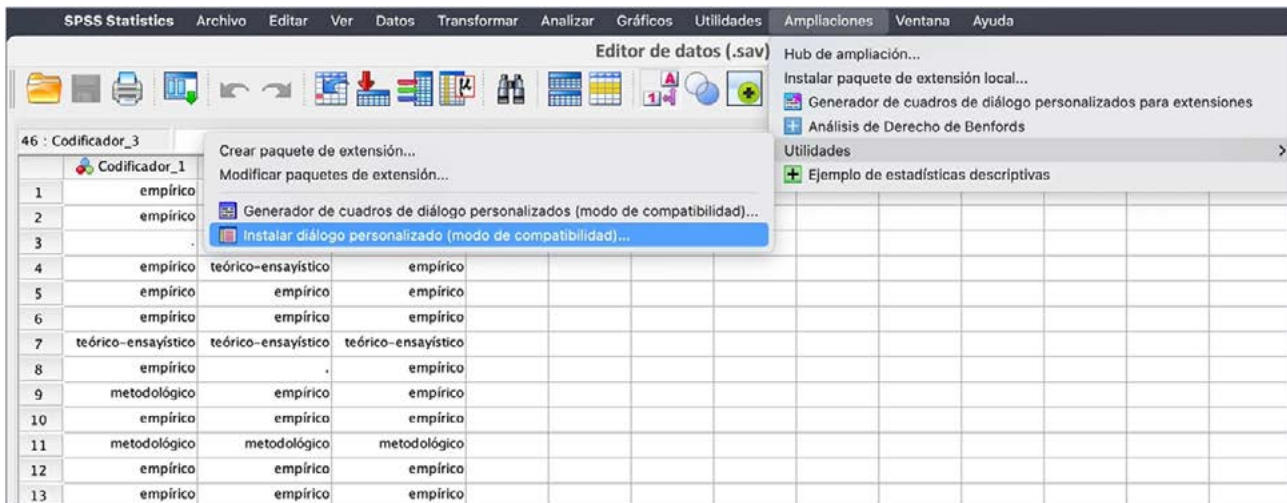
Notemos que el Codificador 1 no reporta su observación en el caso 3, así como tampoco lo hace el Codificador 2 para el caso 8. Hemos introducido valores perdidos de manera premeditada porque, como ya hemos adelantado, el α de Krippendorff es lo suficientemente robusto como para lidiar con esta particularidad.

3.3. Procedimiento de cálculo con interfaz de usuario y sintaxis SPSS

A diferencia de la κ de Cohen, incluida en SPSS, el α de Krippendorff no figura, por defecto, en este *software*. Sin embargo, es posible incorporarla como una extensión mediante la macro KALPHA, desarrollada por Andrew F. Hayes y Klaus Krippendorff y disponible para su descarga en el sitio web del primero de ellos.³

En la sección web “My Macros and Code for SPSS, SAS, and R” existe un apartado (KALPHA) donde consta el enlace a una carpeta comprimida denominada “kalpha.zip”. Clicamos sobre él y se transfiere a nuestro equipo. Si trasladamos la carpeta al escritorio y la descomprimos, tendremos acceso a 5 archivos: 2 documentos explicativos y 3 ejecutables para SPSS, de los que nos interesa, básicamente, el llamado “kalpha.spd” (que está acompañado de un icono circular en color azul). Acto seguido, abrimos el programa (en nuestro caso, la versión 28) y activamos el siguiente procedimiento para añadir la macro: “Menú / Ampliaciones / Utilidades / Instalar diálogo personalizado” (Imagen 14). De inmediato, seleccionamos “kalpha.spd” y aceptamos para culminar el proceso.

Imagen 14. Instalación de la macro KALPHA en SPSS



En el entorno de usuario del SPSS, el Alpha de Krippendorff se encuentra en: “Menú / Analizar / Estadísticos descriptivos / Krippendorff’s_alpha” (Imagen 15). Así pues, y con la matriz de datos ya configurada (a semejanza de la Imagen 13), accedemos a “Krippendorff’s_alpha” y nos aparecerá un cuadro de diálogo donde veremos las variables “Codificador 1”, “Codificador 2” y “Codificador 3”, que arrastraremos a la ventana denominada “Coders”. En el nivel de medida (“Measurement level”) deberemos especificar el tipo de variable que estemos analizando: nominal, ordinal, de intervalo o de razón. En nuestro ejemplo, el tipo de artículo es una variable nominal multicategoría con 3 opciones de respuesta, así que marcaremos “nominal”.

Finalmente, en “Bootstrap samples”, es posible solicitarle a la macro que efectúe hasta 10,000 simulaciones del cálculo de la fiabilidad en función de otras tantas muestras para establecer, así, intervalos de confianza al 95%. En otras palabras, la herramienta ofrecerá dos valores en cuyo rango se encontraría, al 95%, el valor real de α en caso de ser codificado el universo poblacional (y no únicamente una muestra de casos). Por otro lado, cuantas más repeticiones solicitemos a la macro, más tiempo demorará el programa en procesar los datos, pero más fiable será la inferencia estadística (Hayes & Krippendorff, 2007), así que lo más conveniente será “10,000” (Imagen 16).

Imagen 15. Ruta para solicitar la macro KALPHA en SPSS

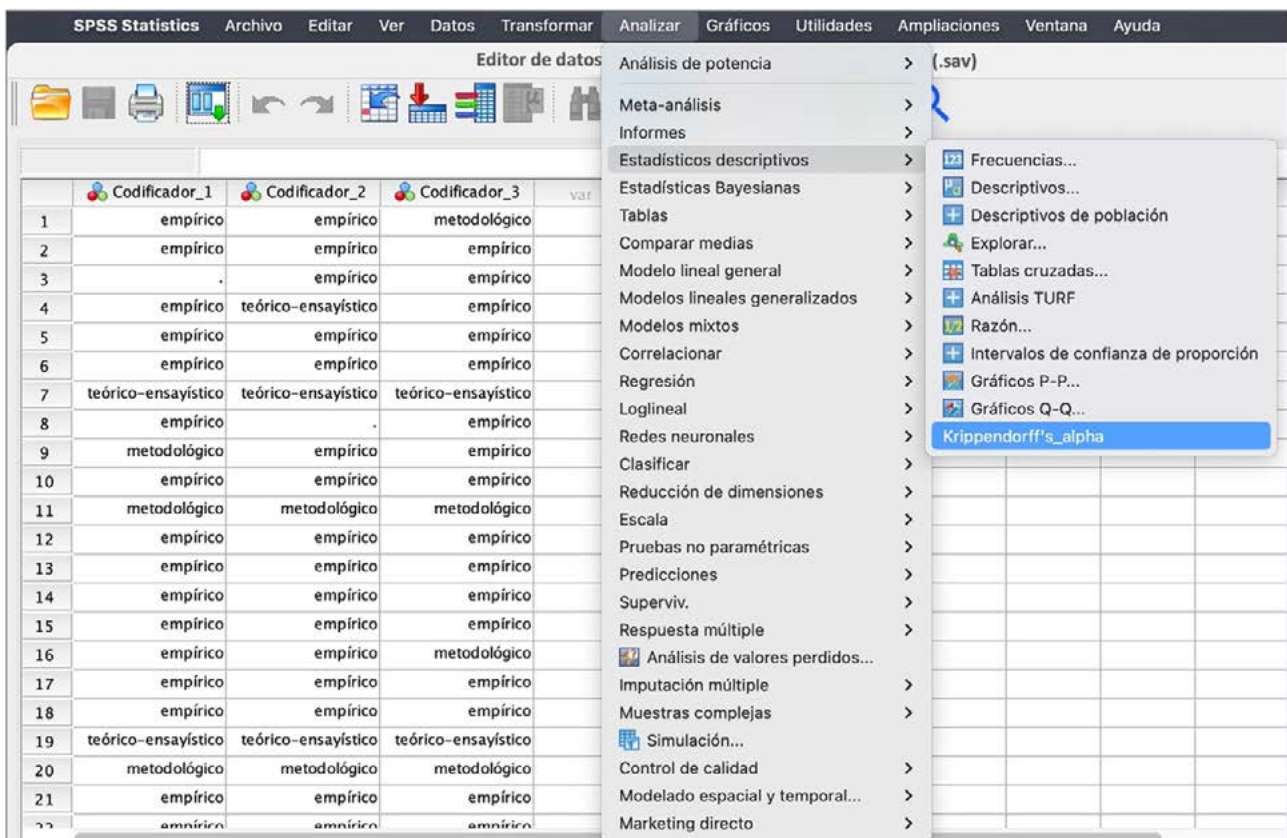
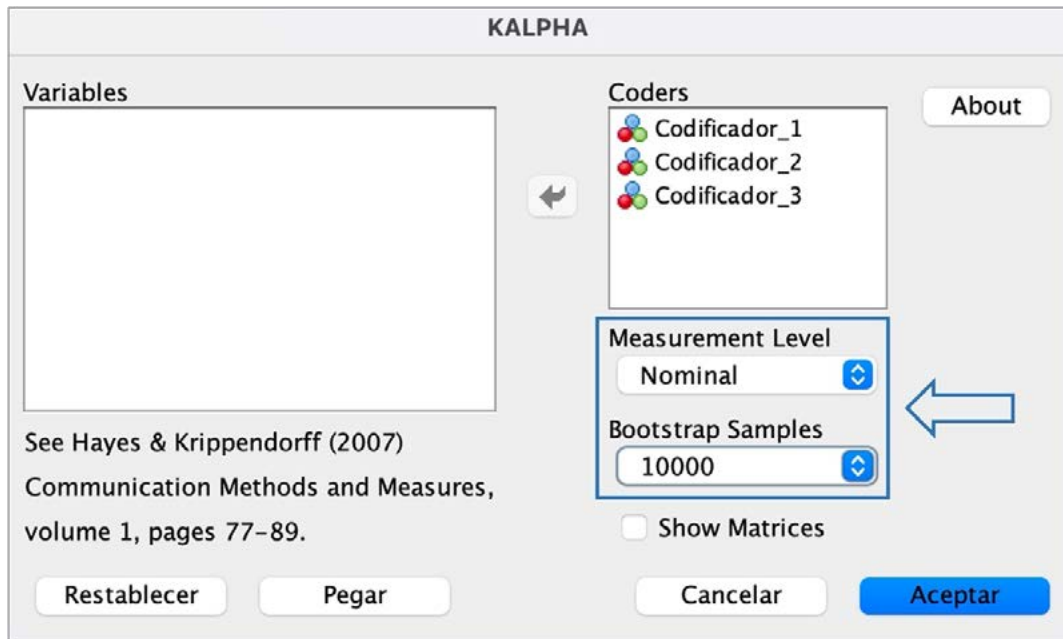


Imagen 16. Ejecución de la macro KALPHA en SPSS



Ya hemos incidido, a propósito de la Kappa de Cohen, en la utilidad de la sintaxis para generar, por así denominarlo, un “cuaderno de bitácora” de la explotación de los datos que sería fácilmente recuperable en caso de que algunos de estos cambiaran. Lo que nunca debería modificarse es el nombre de las variables, sino el programa no será capaz de asociar el comando con los ítems que requiramos para un determinado cruce o procesamiento estadístico (y, si

cambiamos el nombre de las variables en la matriz, habrá que modificarlo también en la sintaxis). En el caso del Alpha de Krippendorff, el comando es bastante sencillo. En el supuesto que nos sirve como referencia, la expresión alfanumérica sería la siguiente (Imagen 17):⁴

```
KALPHA judges = Codificador_1 Codificador_2
Codificador_3/level = 1/detail = 0/boot = 10000
```

Imagen 17. Sintaxis para el cálculo del Alpha de Krippendorff con la macro KALPHA



Como podremos comprobar, la macro se activa en la sintaxis tras el comando “KALPHA”. Seguidamente, y usando un único espacio, escribimos “judges =” para indicar las variables que deseamos cruzar (Codificador_1 Codificador_2 Codificador_3). La siguiente indicación (“/level”) tiene que ver con el nivel de medida de las variables. Por defecto, nominal equivale a “1”, ordinal a “2”, intervalo a “3” y razón a “4”; por lo tanto, introducimos “/level =” seguido del número que corresponda a las variables en cuestión (para nuestro caso, “1”). El subcomando “/detail” le pide a la macro que reporte (valor “1”) u omita (valor “0”) los detalles computacionales derivados de las matrices de coincidencia observadas y esperadas y de la matriz delta (para más detalle, véase:

Krippendorff, 2004a). Finalmente, “/boot” es el encargado de generar 1,000; 2,000; 5,000 o 10,000 muestras. En nuestro caso, omitiremos los detalles computacionales (/detail = 0), pero sí le pediremos a la macro que nos calcule el mayor número posible de simulaciones (/boot = 10,000).

3.4. Resultados

Una vez que activamos todas las opciones anteriormente señaladas en la interfaz de usuario de SPSS (en el fichero “.sav”), o bien ejecutamos el comando especificado en la sintaxis (“.sps”), el programa nos devolverá un archivo de resultados (“.spv”) con los valores concernientes a KALPHA. En el supuesto que nos ocupa, son los que siguen (Imagen 18):

Imagen 18. Captura de pantalla del output que ofrece SPSS al ejecutar la macro KALPHA

```

Run MATRIX procedure:

Krippendorff's Alpha Reliability Estimate

Nominal      Alpha  LL95%CI  UL95%CI  Units  Observrs  Pairs
Nominal      ,7987  ,6577    ,8993    50,0000  3,0000  144,0000

Probability (q) of failure to achieve an alpha of at least alphamin:
alphamin      q
,9000         ,9776
,8000         ,5760
,7000         ,0924
,6700         ,0259
,6000         ,0044
,5000         ,0000

Number of bootstrap samples:
10000

Judges used in these computations:
Coder_1 Coder_2 Coder_3

Examine output for SPSS errors and do not interpret if any are found

----- END MATRIX -----

```

A tenor de la información aportada por la macro (Imagen 18), y tras una breve descripción del procesamiento efectuado (o sea, “*Krippendorff’s Alpha Reliability Estimate*”), la primera fila es la que contiene los datos más relevantes a efectos del chequeo de la fiabilidad. De izquierda a derecha:

- El nivel de medida de las variables: “nominal”.
- El valor del Alpha de Krippendorff: “.7987”.
- El límite inferior del intervalo de confianza al 95%: “.6577”.
- El límite superior del intervalo de confianza al 95%: “.8993”.
- El número de filas, casos o unidades de análisis: “50”.
- El número de jueces, codificadores u observadores: “3”.
- El número de pares de observaciones efectuadas: “144”. Lo recomendable aquí habría sido “150” (50 unidades x 3 observadores), pero al existir 3 celdas con valores perdidos, esa cifra desciende a 144. Concretamente, introdujimos un caso *missing* por cada codificador, de manera que: Codificador 1 vs. 2 y 3, Codificador 2 vs. 1 y 3, y Codificador 3 vs. 1 y 2 suman las 6 observaciones pareadas que KALPHA no ha podido cruzar, de ahí la cifra de “144” (150 – 6).

El siguiente bloque de datos atiende a la probabilidad (q), expresada de 0 a 1 (que podemos transponer a su equivalente porcentual: de 0% a 100%), de obtener un valor mínimo para a fruto del *bootstrapping*, o simulación de 10,000 muestras. Así pues, observamos que $q = .0259$ para $a = .67$; es decir, es altamente probable (estaríamos errando menos de un 3% de las veces) que alcanzásemos una fiabilidad aceptable (.67) si hubiésemos analizado el universo poblacional en lugar de una muestra. Ahora bien, tendríamos muchos problemas ($q = .576$) para alcanzar un Alpha adecuado ($a = .80$) bajo dicho supuesto poblacional. Para terminar, en “*Judges used in these computations*” comprobamos que, en efecto, hemos introducido correctamente las variables que interesaba cruzar (Codificador_1, Codificador_2, Codificador_3).

3.5. Reporte estadístico estandarizado

Exactamente igual que con Kappa, el reporte de Alpha se efectuará en el apartado metodológico del estudio. En líneas generales, primero será preciso aludir a la estrategia de muestreo, luego a las variables que componen el libro de códigos y a la codificación de la muestra y, en última instancia, al chequeo de la fiabilidad intercodificadores. Desde este planteamiento, los resultados del análisis anterior se reportarían así:

“Se calcula el Alpha de Krippendorff (Krippendorff, 2011) con el objetivo de medir la fiabilidad entre los 3 codificadores encargados de realizar el análisis de contenido. En concreto, estos analistas han inferido, con arreglo a una muestra de 50 artículos, cuál era su tipología: empíricos, teórico-ensayísticos o metodológicos. Para determinar la fiabilidad entre los 3 codificadores controlando el azar, se calcula el Alpha de Krippendorff mediante la macro KALPHA para SPSS (Hayes & Krippendorff, 2007), que arroja un coeficiente de: $a = .798$ (IC 95%: .657, .899, *bootstrap* = 10,000), evidenciando una adecuada fiabilidad entre codificadores”.

4. Consideraciones finales

A lo largo de este trabajo hemos tratado de explicar, en detalle, las suposiciones estadísticas y el procedimiento de cómputo de dos pruebas esenciales para medir la fiabilidad entre codificadores en un análisis de contenido: la Kappa de Cohen y el Alpha de Krippendorff. Para ello, hemos utilizado como paquete estadístico un *software* muy popular entre la comunidad investigadora en ciencias sociales (SPSS), tomando como ejemplo un análisis de contenido consistente en la codificación de variables procedentes de un artículo científico (una práctica conocida como “meta-investigación”).

De las dos pruebas estadísticas, hemos destacado el Alpha de Krippendorff como el coeficiente más robusto para el análisis de fiabilidad entre codificadores debido a sus propiedades, pues permite efectuar la recogida de datos con más de dos codificadores,

lidia con posibles casos vacíos y funciona con variables ordinales o de razón, aspectos que la Kappa de Cohen no cubre. Sin embargo, esta es una prueba muy aceptada, simple y suficiente para la ejecución de la mayoría de análisis de contenido estándar en el campo, una faceta que permite su generalizada utilización. Esta carta de presentación de la Kappa de Cohen se complementa, además, con su inclusión en la mayoría de paquetes estadísticos para análisis de datos, incluido SPSS, lo que la convierte en un test sobradamente conocido y usado. Por el contrario, el Alpha de Krippendorff, a pesar de sus mejores prestaciones, no está presente en los paquetes estadísticos más populares, lo que dificulta su generalización y ejecución por la comunidad interesada. Esta ausencia, sin embargo, ha sido paliada gracias a la utilización de macros como KALPHA, desarrollada por Andrew F. Hayes y el propio Klaus Krippendorff y abordada en esta guía.

Como explicamos a lo largo del trabajo, la incorporación de un coeficiente que establezca la fiabilidad de la codificación es esencial para la reproductibilidad de los análisis y la solvencia de las inferencias estadísticas que se deriven de los datos, en cualquier disciplina científica. De hecho, la ausencia (o deficiencia) de dicho coeficiente sería condición suficiente para cuestionar los hallazgos encontrados y/o poder rechazar un estudio. Así las cosas, queda claro que la fiabilidad se erige en una fase crucial del protocolo de aplicación de todo análisis de contenido porque revela, en gran medida, la transparencia y la objetividad con que se ha efectuado la codificación.

Para finalizar, conviene asimismo destacar que la combinación de métodos computacionales en el análisis de contenido representa, hoy en día, un avance significativo para la exploración de grandes bases de datos. Tanto es así que la codificación asistida por ordenador como las técnicas más avanzadas de *machine learning* han permitido análisis cada vez más eficientes, manteniendo la fiabilidad. Sin embargo, a pesar de estas innovaciones metodológicas y su aplicación en la investigación, la importancia del procedimiento descrito en el presente artículo es fundamental por varias razones.

En primer lugar, porque el procedimiento descrito en el presente estudio puede servir como punto de referencia comparativo con el que evaluar la fiabilidad de las nuevas metodologías computacionales y los resultados obtenidos con ellas. En segundo lugar, porque el procedimiento descrito puede ser más accesible y, por lo tanto, más fácil de implementar por investigadores que no tengan acceso a recursos tecnológicos avanzados o que no tengan conocimientos técnicos para el diseño y ejecución de análisis con metodologías computacionales sofisticadas. En tercer y último lugar, la automatización mediante metodologías computacionales permite, sin duda, análisis de contenido más eficientes y extensos, pero no es menos cierto que en muchos casos el entrenamiento de algoritmos de clasificación exige previamente la interpretación y análisis crítico de la inteligencia humana, algo que, hasta la fecha, la computación inteligente todavía no ha conseguido substituir.

5. Notas

¹ Los archivos de SPSS que se han usado para la presente guía práctica, denominados: “matriz_datos_k_a.sav”, “sintaxis_k_a.sps” y “resultados_k_a.spv”, están disponibles para su descarga en: <https://doi.org/10.17605/OSF.IO/KNW2G>.

² Para realizar las capturas de pantalla que sirven como ilustración de los distintos procesos detallados se ha utilizado la versión 28 del programa SPSS para “macOS”.

³ El enlace para efectuar la descarga de la macro KALPHA para SPSS es el siguiente: <https://www.afhayes.com/spss-sas-and-r-macros-and-code.html>.

⁴ Si el comando de sintaxis para KALPHA no funcionase en primera instancia, bastaría con ejecutar una única vez el estadístico mediante la interfaz de usuario y, a partir de las siguientes ocasiones ya se podría emplear el código.

6. Referencias

- Altman, D. G. (1999). *Practical statistics for medical research*. Chapman & Hall.
- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational linguistics*, 34(4), 555-596. <https://doi.org/10.1162/coli.07-034-R2>
- Bardin, L. (1986). *Análisis de contenido*. Akal.
- Bennett, E. M., Alpert, R., & Goldstein, A.C. (1954). Communications through limited-response questioning. *Public opinion quarterly*, 18(3), 303-308. <https://doi.org/10.1086/266520>
- Berelson, B. (1952). *Content analysis in communication research*. Free Press.
- Berger, A. A. (2016). *Media and communication research methods. An introduction to qualitative and quantitative approaches (4th Ed.)*. Sage.
- Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International journal of educational research*, 31(8), 659-671. [https://doi.org/10.1016/S0883-0355\(99\)00032-4](https://doi.org/10.1016/S0883-0355(99)00032-4)
- Caffarel-Serra, C., Ortega-Mohedano, F., & Gaitán-Moya, J. A. (2017). Investigación en comunicación en la universidad española en el periodo 2007-2014. *Profesional de la información*, 26(2), 218-227. <https://doi.org/10.3145/epi.2017.mar.08>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5), 378-382. <https://doi.org/10.1037/h0031619>
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions (3rd Ed.)*. John Wiley & Sons.
- Freelon, D. (2013). ReCal OIR: ordinal, interval, and ratio intercoder reliability as a web service. *International journal of internet science*, 8(1), 10-16.
- Goyanes, M., Rodríguez-Gómez, E. F., & Rosique-Cedillo, G. (2018). Investigación en comunicación en revistas científicas en España (2005-2015): de disquisiciones teóricas a investigación basada en evidencias. *Profesional de la Información*, 27(6), 1281-1291. <https://doi.org/10.3145/epi.2018.nov.11>

- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89. <https://doi.org/10.1080/19312450709336664>
- Igartua, J. J. (2006). *Métodos cuantitativos de investigación en comunicación*. Bosch.
- Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: an examination of applications with directives for improving research reliability and objectivity. *Journal of consumer research*, 18(2), 243-250. <https://doi.org/10.1086/209256>
- Krippendorff, K. (2004a). *Content analysis: an introduction to its methodology (2nd Ed.)*. Sage.
- Krippendorff, K. (2004b). Reliability in content analysis. Some common misconceptions and recommendations. *Human communication research*, 30(3), 411-433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication methods and measures*, 5(2), 93-112. <http://dx.doi.org/10.1080/19312458.2011.568376>
- Lacy, S., Watson, B. R., Riffe, D., & Lovejoy, J. (2015). Issues and best practices in content analysis. *Journalism & mass communication quarterly*, 92(4), 791-811. <https://doi.org/10.1177/1077699015607338>
- Lombard, M., Snyder-Duch, J., & Campanella, C. (2002). Content analysis in mass communication. Assessment and reporting of intercoder reliability. *Human communication research*, 28(4), 587-604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- Lovejoy, J., Watson, B. R., Lacy, S. & Riffe, D. (2014). Assessing the reporting of reliability in published content analyses: 1985-2010. *Communication methods and measures*, 8(3), 207-221. <https://doi.org/10.1080/19312458.2014.937528>
- Lovejoy, J., Watson, B. R., Lacy, S., & Riffe, D. (2016). Three decades of reliability in communication content analyses: reporting of reliability statistics and coefficient levels in three top journals. *Journalism & mass communication quarterly*, 93(4), 1135-1159. <https://doi.org/10.1177/1077699016644558>
- Manterola, C., Grande, L., Otzen, T., García, N., Salazar, P., & Quiroz, G. (2018). Confiabilidad, precisión o reproducibilidad de las mediciones. Métodos de valoración, utilidad y aplicaciones en la práctica clínica. *Revista chilena de infectología*, 35(6), 680-688. <https://dx.doi.org/10.4067/S0716-10182018000600680>
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Sage.
- Neuendorf, K. A. & Kumar, A. (2016). Content analysis. En G. Mazzoleni (ed.), *The International Encyclopedia of Political Communication*. Wiley. <https://doi.org/10.1002/9781118541555.wbiepc065>
- Piñeiro-Naval, V. (2020). The content analysis methodology. Uses and applications in communication research on Spanish-speaking countries. *Communication & society*, 33(3), 1-15. <http://dx.doi.org/10.15581/003.33.3.1-15>
- Piñeiro-Naval, V., & Morais, R. (2019). Estudio de la producción académica sobre comunicación en España e Hispanoamérica. *Comunicar*, 27(61), 113-123. <http://dx.doi.org/10.3916/C61-2019-10>
- Piñuel, J. L. (2002). Epistemología, metodología y técnicas del análisis de contenido. *Estudios de sociolingüística*, 3(1), 1-42.
- Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of applied communication research*, 27(3), 258-284. <https://doi.org/10.1080/00909889909365539>
- Riffe, D., & Freitag, A. (1997). A content analysis of content analyses: twenty-five years of journalism quarterly. *Journalism & mass communication quarterly*, 74(3), 515-524. <https://doi.org/10.1177/107769909707400306>
- Riffe, D., Lacy, S., Watson, B. R., & Fico, F. (2019). *Analyzing Media Messages. Using Quantitative Content Analysis in Research (4th Ed.)*. Routledge.
- Scott, W. A. (1955). Reliability of content analysis: the case of nominal scale coding. *Public opinion quarterly*, 19(3), 321-325. <https://doi.org/10.1086/266577>
- Segado-Boj, F., Piñeiro-Naval, V., & Antona-Jimeno, T. (2023). Spanish research on communication in WoS: thematic, methodological, and intellectual comparison between SSCI and ESCI. *Profesional de la información*, 32(3), e320309. <https://doi.org/10.3145/epi.2023.may09>
- Singletary, M. W. (1993). *Mass communication research: contemporary methods and applications*. Addison-Wesley.
- Stemler, S. E. (2015). Content analysis. En: Robert A. Scott; Marlis C. Buchmann (eds.), *Emerging trends in the social and behavioral sciences*. Wiley. <https://doi.org/10.1002/9781118900772.etrds0053>
- Walter, N., Cody, M. J., & Ball-Rokeach, S. J. (2018). The ebb and flow of communication research: seven decades of publication trends and research priorities. *Journal of communication*, 68(2), 424-440. <https://doi.org/10.1093/joc/jqx015>
- Wimmer, R. D., & Dominick, J. R. (2011). *Mass media research: an introduction (9th Ed.)*. Wadsworth.

Manuel Goyanes es Profesor e Investigador en Metodologías de Investigación en la Universidad Carlos III de Madrid. Su investigación aborda la influencia del periodismo y las nuevas tecnologías en la vida cotidiana de los ciudadanos, así como los efectos del consumo de noticias en el conocimiento y la participación política. También ha estudiado las desigualdades globales en la participación académica, los sesgos sistemáticos hacia los académicos del Sur global y las tendencias de publicación en Comunicación. Sus trabajos han sido publicados en revistas como *Communication Research*, *Plos One*, *New Media & Society*, etc. ORCID: <https://orcid.org/0000-0001-8329-0610>

Valeriano Piñeiro-Naval es Profesor Permanente Laboral en el Departamento de Sociología y Comunicación de la Universidad de Salamanca, así como miembro investigador del Observatorio de los Contenidos Audiovisuales en esa misma institución. Su investigación gira en torno a las dinámicas que se establecen entre la cultura, el patrimonio y el turismo en la sociedad digital, el análisis de contenido y diseño de sitios web municipales, las narrativas transmedia y la metainvestigación en ciencias de la comunicación. Sus trabajos han sido publicados en revistas como *Comunicar*, *Revista Española de Documentación Científica*, *Communication & Society*, etc. ORCID: <https://orcid.org/0000-0001-9521-3364>