



## Correlatie

Sil Aarts · Anja de Koning

Published online: 27 November 2018

© Bohn Stafleu van Loghum is een imprint van Springer Media B.V., onderdeel van Springer Nature 2018

**Samenvatting** Als laatste analysemethode van deze statistiekreeks, zal de correlatie besproken worden. Deze statistische maat is wellicht het meest bekend bij medisch onderzoekers. Correlatie is een maat om het verband tussen twee, vaak numerieke, variabelen weer te geven. Het meest gebruikte voorbeeld betreft ‘het verband tussen lengte en gewicht’: het gewicht correleert bijvoorbeeld met de lengte, dat wil zeggen dat iemands gewicht in het algemeen toeneemt wanneer mensen langer zijn.

**Trefwoorden** correlatie · scatterplot · Pearson-correlatie · Spearman-correlatie

### Inleiding

Correlatie bepaalt in hoeverre twee variabelen ‘proportioneel’ ten opzichte van elkaar een lineaire afhankelijkheid vertonen, dat wil zeggen, met elkaar samenhangen. Correlatie varieert van ‘geen lineaire relatie’ ( $r=0$ ) tot een ‘perfect lineaire relatie’ (zie kader Scatterplot). Bij een perfect lineaire relatie kan er sprake zijn van respectievelijk een perfect positief lineaire relatie ( $r=1$ ) of een perfect negatieve lineaire relatie ( $r=-1$ ). Over het algemeen geldt dat, rekening houdend met de context van het betreffende onderzoek, een waarde van 0,2 of minder als een zwakke

samenhang wordt beschouwd, een waarde tussen 0,2 en 0,5 als een gemiddelde samenhang en een waarde van 0,8 of groter als een sterke samenhang [1, 2].

De correlatie wordt veelal aangegeven met de letter  $r$  of de Griekse letter *rho* ( $\rho$ ). De hypothesen die bij een correlatie worden gesteld zijn:

**H<sub>0</sub>** Er is geen sprake van correlatie;  $\rho = 0$

**H<sub>1</sub>** Er is sprake van correlatie;  $\rho \neq 0$

Om te testen of twee variabelen met elkaar samenhangen, kunnen met behulp van statistische programma's verschillende correlatiecoëfficiënten worden berekend, waarvan de Pearson-correlatiecoëfficiënt de meest voorkomende is [1]. De Pearson-correlatie wordt gebruikt wanneer beide variabelen numeriek zijn (dat wil zeggen variabelen waarmee kan worden gerekend; waarbij bijvoorbeeld geldt, 5 is  $2 \times$  meer dan 10, en het verschil tussen 7 en 9 is hetzelfde verschil als het verschil tussen 8 en 10). Als ten minste een van de twee variabelen van ordinaal meetniveau is (bijvoorbeeld slecht, matig, voldoende, goed, uitste-

### Casus

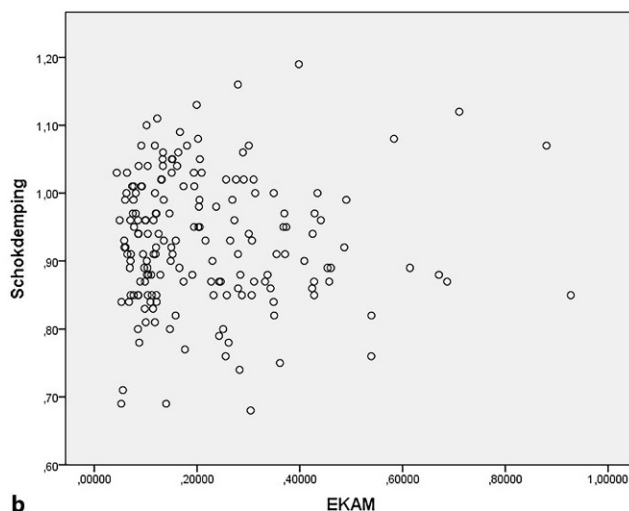
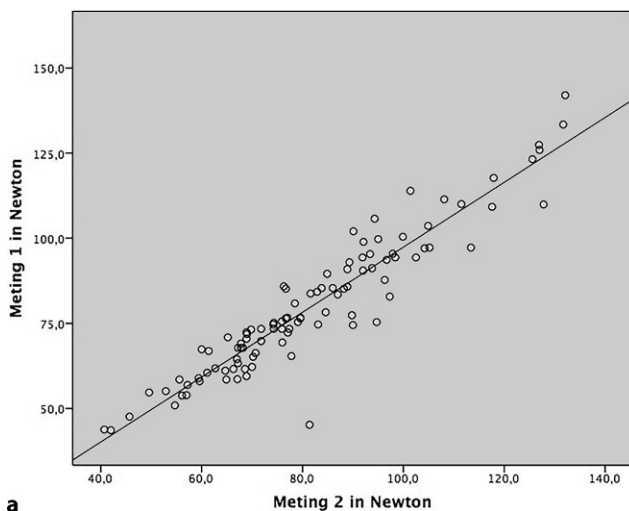
Een onderzoeksvraag die in de podotherapeutische praktijk zou kunnen worden gesteld en met behulp van een correlatieanalyse zou kunnen worden beantwoord, is: ‘Wat is het verband tussen het aantal behandelingen dat een patiënt heeft ondergaan en zijn tevredenheidsscore over deze behandeling?’. Deze onderzoeksvraag bevat twee numerieke variabelen: ‘het aantal behandelingen’ en ‘de tevredenheidsscore’. Deze laatste kan bijvoorbeeld in kaart worden gebracht door een score op een schaal van 1 tot 100 te geven.

In deze rubriek dragen de auteurs een steentje bij aan het vergroten van de kennis over wetenschappelijk onderzoek en de toepasbaarheid ervan in de podotherapeutische praktijk.

dr. S. Aarts (✉)  
 Health Services Research, Universiteit Maastricht,  
 Maastricht, Nederland  
[s.aarts@maastrichtuniversity.nl](mailto:s.aarts@maastrichtuniversity.nl)

A. de Koning  
 VUmc, Amsterdam AMC, Amsterdam, Nederland





**Figuur 1** Scatterplots. **a** Een positief sterke correlatie met een regressielijn erdoor getrokken en **b** een lage, waarschijnlijk, negatieve correlatie met outliers.

(De scatterplots zijn afkomstig uit afstudeerprojecten van Ayra Bekker-de Waard en Wouter van Diessen aan de Fontys Paramedische Hogeschool)

kend), kan de Spearman-correlatie gebruikt worden. Deze correlatie is gebaseerd op rangordes (voor meer informatie, zie de referenties); van de scores van de twee betrokken variabelen wordt de rang bepaald en deze rangen worden gebruikt om de correlatie te berekenen.

**Voorwaarden**

Er moet aan drie voorwaarden zijn voldaan om een correlatie te kunnen bepalen. Als aan ten minste een van de volgende voorwaarden niet wordt voldaan, is een correlatie mogelijk niet te gebruiken.

1. *Er moet sprake zijn van een random steekproef en de observaties dienen onafhankelijk te zijn.* Dit betekent dat de waarnemingen aselekt moeten worden getrokken (ofwel, elke persoon heeft een even grote kans om aan het onderzoek deel te nemen). Het betekent ook dat de waarnemingen onafhankelijk moeten zijn, wat in ons voorbeeld inhoudt dat patiënten zich niet door andere patiënten laten beïnvloeden in hun tevredenheid over de behandeling en dat elke patiënt slechts een keer mag deelnemen aan dit onderzoek.
2. *Het verband moet lineair zijn zonder enorme outliers.* Een lineair verband houdt in dat er sprake is van gelijke toename of afname. Er mag bijvoorbeeld

geen exponentieel verband zijn (dan is er sprake van een onevenredige toe- of afname). Met behulp van een scatterplot kan worden gecontroleerd of er outliers zijn, ook wel uitschieters of uitbijters genoemd (zie kader Scatterplot).

3. *De steekproef moet groot genoeg zijn.* Als vuistregel wordt gesteld dat er ten minste sprake moet zijn van 30-waarnemingen of observaties.

**Correlatietabel**

Wanneer er een statistisch programma wordt gebruikt om te onderzoeken in hoeverre er een correlatie bestaat tussen variabelen, kan er een correlatietabel worden gegenereerd, bijvoorbeeld tab. 1 [3]. In deze fictieve tabel zijn twee numerieke variabelen te vinden: lengte en gewicht. De onderzoeksvraag die hierbij geformuleerd kan worden is: ‘Wat is de relatie tussen lengte en gewicht?’. Of, met andere woorden: ‘Wat is de correlatie/samenhang tussen lengte en gewicht?’. De *Pearson-correlatie* geeft de correlatie aan. Zoals eerder genoemd, kan deze correlatie een waarde aannemen tussen de  $-1$  en  $+1$ . Tab. 1 laat twee correlatiewaarden zien, respectievelijk 1 en 0,070. De eerste waarde bevat de correlatie tussen lengte en lengte; aangezien dit de correlatie met de variabele zelf is, is deze altijd 1. De correlatie van 0,070 is de samenhang tussen de variabelen lengte en gewicht.

**Tabel 1** Correlatietabel uit SPSS

		lengte	gewicht
lengte	Pearson Correlation	1	0,070
	Sig. (2-tailed)		0,621
	N	52	52
gewicht	Pearson Correlation	0,070	1
	Sig. (2-tailed)	0,621	
	N	52	52



### Scatterplot

Om inzichtelijk te maken in hoeverre de waarden van twee numerieke variabelen met elkaar samenhangen, kan men een scatterplot maken. Hierin worden de waarden tegen elkaar afgezet in een grafiek (zie fig. 1). Elk punt in een dergelijke plot stelt één deelnemer voor; hoe meer puntjes de plot bevat, hoe meer deelnemers. Als er sprake is van een sterke correlatie, is het mogelijk om een ellips rond de punten te tekenen, die kan worden benaderd door een opwaartse of een neerwaartse rechte lijn (fig. 1a). Wanneer er sprake is van een zwakke samenhang, is dit niet mogelijk, omdat dan alle punten verspreid ('scattered') zijn (fig. 1b). *Outliers* of uitschieters zijn punten die niet lijken te passen bij de rest van de punten en die zich voornamelijk aan de randen van de grafiek bevinden (bijvoorbeeld rechts in de bovenhoek of links in de benedenhoek).

Zoals te zien aan de  $r$ -waarde, is deze correlatie zwak te noemen. De *Sig. (2-tailed)* staat voor de  $p$ -waarde. Een  $p$ -waarde van  $<0,05$  wordt gezien als statistisch significant (in het volgende en laatste artikel bespreken we de arbitraire indeling van  $p$ -waarden). De letter  $N$  staat voor het aantal deelnemers waarvan de data zijn gebruikt voor deze analyse.

### Kanttekening bij een correlatietoets

Een statistisch significant verband tussen twee variabelen impliceert niet noodzakelijkerwijs een sterke correlatie. Als een steekproefomvang bijvoorbeeld 50 is (dat wil zeggen dat 50 personen deelnemen aan een wetenschappelijk onderzoek), blijken Pearson-correlaties van 0,28 of hoger dikwijls statistisch significant

te zijn (ofwel een  $p$ -waarde  $\leq 0,05$ ), terwijl een correlatie van 0,30 zelden (dit hangt natuurlijk ook af van de context van het onderzoek) als een sterke associatie wordt beschouwd. Dit impliceert dat, hoewel een waargenomen correlatie statistisch significant zou kunnen zijn, deze nog steeds triviaal kan zijn als er sprake is van een kleine correlatie. Het is bij het bepalen van de klinische relevantie van de uitkomsten van een correlatietoets dan ook belangrijk dat er niet alleen gefocust wordt op de  $p$ -waarde [2].

### Take Home Message

Ook in de podotherapeutische praktijk kan de correlatie een meerwaarde hebben bij het onderzoek naar de samenhang van variabelen. Let wel, als er een statistisch significante samenhang wordt gevonden, betekent dat niet direct dat er een causaal – dat wil zeggen oorzaak-gevolg – verband bestaat. Als bijvoorbeeld uit correlatieonderzoek blijkt dat de tevredenheid samenhangt ('correleert') met het aantal behandelingen, wordt de hogere tevredenheid niet per se veroorzaakt door het aantal behandelingen. Causaliteit kan dan ook niet uit een correlatieanalyse worden verondersteld.

### Literatuur

1. Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. *Am Stat.* 1988;42(1):59–66.
2. Cohen J. The earth is round ( $p < .05$ ). *Am Psychol.* 2004;49(12):997–1003.
3. Field A. *Discovering statistics using SPSS*. 3e druk. Londen: SAGE; 2009.

**dr. Sil Aarts**, universitair docent

**Anja de Koning**, beleidsmedewerker Faculteit der Gedrags- en Bewegingswetenschappen