

Hoofdstuk 4. Nominale Associatiematen

Bij centrum- en spreidingsmaten is het van belang om te weten wat het meetniveau is om te bepalen welke analysemogelijkheden er zijn. Hetzelfde is van toepassing bij associatiematen. Dit hoofdstuk behandelt associatiematen op nominaal niveau zoals *Cramers V*, *phi*, *Goodman & Kruskals tau* en *lambda*.

Associatiematen

Een verband tussen twee variabelen kan worden aangetoond met een associatiemaat. In plaats van associatiematen kan ook gesproken worden van samenhangsmaten. Bij de sterkte van een verband gaat het om nominale variabelen. Bij sterkte en richting van een verband spreek je over variabelen met overige meetniveaus. Een voorbeeld: je wilt nagaan of er een sterk of zwak verband is tussen iemands politieke voorkeur en welke televisie programma's men kijkt. De associatiemaat kan aangeven of hierbij een verschil optreedt. De sterkte van een verband op nominaal niveau wordt weergegeven in cijfers tussen '0' → geen verband en '1' → sterk verband. Deze associatiemaat kan nooit negatief zijn, omdat de richting niet kan worden aangetoond.

Meetniveau

Op basis van het meetniveau kun je bepalen welke associatiemaat je moet gebruiken. Ook al is maar één van de variabelen nominaal, dan gebruik je toch een associatiemaat op nominaal niveau.

Symmetrie en asymmetrie

Eerst bepaal je dus het meetniveau en daarna kijk je naar de veronderstelde relatie tussen twee variabelen. Bij twee variabelen kan de ene onafhankelijk zijn en de ander daarvan afhankelijk. Voorbeeld: bij een onderzoek veronderstel je, ondersteund door de gevonden theorie, dat iemands partijvoorkeur afhangt van het soort televisieprogramma dat iemand kijkt. Partijvoorkeur is dan de afhankelijke variabele, want die is afhankelijk van het soort tv-programma dat iemand kijkt (onafhankelijke variabele). Hier spreekt men van een **asymmetrische** relatie. Kun je een dergelijk onderscheid niet onderbouwen en is de richting van de beïnvloeding van de variabele niet duidelijk, dan is de relatie **symmetrisch**.

Er zijn ook variabelen die per definitie onbeïnvloedbaar zijn en dus onafhankelijk zoals; geslacht of leeftijd. Als je niet op de hoogte bent van de onafhankelijke en afhankelijke variabele is het niet mogelijk om een asymmetrische associatiemaat uit te rekenen, wel kun je dan een symmetrische associatiemaat berekenen.

Interpreteren

Bij het interpreteren van de associatiematen kun je de volgende grove richtlijnen gebruiken:

0	-	0,10:	Zeer zwak/geen verband
0,11	-	0,30:	Zwak verband
0,31	-	0,50:	Redelijk verband
0,51	-	0,80:	Sterk verband
0,81	-	0,99:	Zeer sterk verband
		1:	Perfect verband

Vanaf ordinaal meetniveau is het ook van belang om de richting van het verband aan te geven: positief (+) en negatief (-).

JoHo Samenvatting – Beschrijvende Statistiek

Cramers V

Cramers V is een associatiemaat die je berekent bij een nominale variabele en een symmetrische relatie. Aan de hand van Cramers V kun je zien hoe sterk dit verband is. Ligt de uitkomst dicht bij nul, dan is er sprake van een zwak verband. Ligt het dicht bij één, dan is het een sterk verband. Je kunt een verband aflezen uit een kruistabel aan de hand van de percentages, waarbij gepercenteerd is op de kolommen. Ook kun je SPSS vragen om Cramers V te berekenen. Dit doe je door in SPSS naar Analyse → Descriptive Statistics → Crosstabs te gaan. Je geeft bij Cells aan om te percenteren op de kolommen. In communicatiewetenschap wordt doorgaans gebruik gemaakt van kolompercentages. De onafhankelijke variabele wordt daarom in de kolommen gezet. Daarna klik je op Statistics en hier geef je de associatiemaat naar keuze aan.

Berekening

Formule voor Cramers V:

$$\sqrt{\frac{\chi^2}{\chi^2 \text{ maximaal}}} = \sqrt{\frac{\chi^2}{n [\min(r,k) - 1]}}$$

χ^2 maximaal is de maximale waarde die χ^2 (chi-kwadraat) kan aannemen, dat is dus $n_{[\min(r,k) - 1]}$. Min r,k is het kleinste getal van een aantal rijen of kolommen, van dit getal wordt altijd 1 afgetrokken en daarna vermenigvuldigd met het totale aantal onderzoekseenheden. Als laatste wordt de wortel getrokken.

Cramers V is opgebouwd uit χ^2 (chi-kwadraat). χ^2 geeft de sterkte aan van het verband tussen de variabelen, maar is niet handig om te gebruiken, omdat hij niet naar boven is begrensd. Formule voor χ^2 :

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Wat belangrijk is bij de berekening van Cramers V, is het verschil tussen de werkelijk gevonden tabel frequenties (f_o) en de verwachte celfrequenties die je verwacht als er geen verband is tussen beide variabelen (f_e). Hierbij staat f_o voor frequencies observed, dit zijn absolute aantallen in de cellen van een kruistabel. Daarnaast staat f_e voor frequencies expected, de aantallen die je verwacht als je uitgaat van de randtotalen.

Wanneer voor alle cellen de geobserveerde en de verwachte frequenties aan elkaar gelijk zijn ($f_o = f_e$), dan is er geen verband want $\chi^2 = 0$. De meest rechtse kolom en de onderste rij in de kruistabel (randtotalen en percentages) zijn zeer belangrijk bij de berekening van de f_e 's. χ^2 wordt berekend door per cel het verschil tussen f_e en f_o te kwadrateren en dan te delen door f_e . Als laatste tel je alle uitkomsten bij elkaar op.

JoHo Samenvatting – Beschrijvende Statistiek

Phi

Phi is een symmetrische associatiemaat bij variabelen op nominaal niveau die alleen wordt gebruikt bij 2x2 tabellen. Daarbij heeft phi dezelfde waarde aan als Cramers V. Bij de berekening van phi staat ook χ^2 centraal. De formule van phi:

$$\text{Phi} = \sqrt{\frac{\chi^2}{n}}$$

χ^2 wordt op dezelfde manier berekend als in de vorige paragraaf [(fo - fe)² / fe, per cel berekenen en daarna alles optellen].

Goodman & Kruskals tau

Bij deze associatiemaat gaat het om een asymmetrische relatie, dat betekent dat er dus een onafhankelijke en een afhankelijke variabele is. Het gaat hierbij om de voorspelbaarheid van de afhankelijke variabele.

De formule van Goodman & Kruskals tau:

$$\text{Goodman \& Kruskals tau} = \frac{(E_1 - E_2)}{E_1}$$

Hiermee wordt de proportionele voorspellingsverbetering aangegeven van y, rekening houdend met x. E₁ is het aantal voorspellingsfouten *zonder* informatie van x en E₂ is het aantal voorspellingsfouten *met* informatie van x. Hoe beter de voorspelling, hoe kleiner de fout.

Bij een goede voorspelling heeft tau de waarde van 1, wanneer de voorspelling niet zo goed verbeterd kan worden, heeft tau de waarde 0. Voorspellingskans houdt in dat je de kans berekent in hoeveel gevallen je voorspelling juist is en in welke gevallen niet.

Goodman & Kruskals tau kun je ook door SPSS laten berekenen, dit kan bijna op dezelfde manier als bij Cramers V (Analyse → Descriptive Statistics → Crosstabs, dan in Cells kies je Statistics en vervolgens geef je de juiste associatiemaat aan). In het geval van Goodman & Kruskals tau zit deze verstopt achter Lambda, op deze moet je dus klikken om (ook) tau te laten berekenen. Er zijn twee opties, namelijk de 'onafhankelijke variabele-dependent' en de 'afhankelijke variabele-dependent'. SPSS weet namelijk niet welke variabele wij als onafhankelijk hebben gebruikt en welke als afhankelijk. Je kijkt dus naar de 'afhankelijke variabele-dependent'. Kortom, met Goodman & Kruskals tau kun je berekenen hoeveel fouten je maakt bij je voorspellingen als je rekening houdt met de onafhankelijke variabele.

JoHo Samenvatting – Beschrijvende Statistiek

Lambda

Lambda (λ) is een nominale associatiemaat die bruikbaar is wanneer er een afhankelijke variabele is. Lambda is gebaseerd op voorspellingsverbetering. Het verschil met Goodman & Kruskals Tau is dat de voorspellingsfouten worden berekend door gebruik te maken van de modus. De modus is de meest voorkomende waarde oftewel de hoogste frequentie.

Lambda is gebaseerd op dezelfde formule als Goodman & Kruskals Tau:

$$\text{Lambda } (\lambda) = \frac{(E_1 - E_2)}{E_1}$$

E_1 is $n - f_{Mo}(y)$ oftewel het aantal onderzoekseenheden (n) verminderd met de frequentie (f) van de modus (Mo) van y . Hierbij maak je dus alleen gebruik van de afhankelijke variabele.

E_2 is $n - \sum f_{Mo}(y)_{kx}$ oftewel het aantal onderzoekseenheden (n) verminderd met de optelsom (\sum) van de frequentie (f) van de modus (Mo) van y bij gebruikmaking van de informatie over x .

Samenvattend overzicht

Associatiematen geven de sterkte aan van de samenhang tussen twee variabelen en of er überhaupt een verband bestaat tussen twee variabelen. Als één van de variabelen zich op nominaal niveau bevindt en er is sprake van een **symmetrische** relatie, dan kies je altijd voor **Cramers V**. Alleen wanneer het een 2x2 tabel is, kies je voor **phi**. Als de relatie **asymmetrisch** is, dan kies je voor **Goodman & Kruskals tau** of **lambda**.