

Hoofdstuk 8. Eta associatiemaat

Eta en eta-kwadraat

Tot nu toe hebben we steeds gekeken naar het meetniveau van de variabelen om een keuze te maken uit de diverse associatiematen. Normaalgesproken is het zo dat wanneer twee variabelen verschillende meetniveaus hebben, we alleen de associatiematen kunnen gebruiken die bij het laagste niveau horen. Hierop is een uitzondering, namelijk de associatiemaat eta. Deze wordt gebruikt wanneer de afhankelijke variabele interval- of rationiveau is en de onafhankelijke variabele nominaal is, zoals geslacht (man/vrouw). We hoeven ons echter niet te beperken tot associatiematen van nominaal niveau in deze situatie.

Interpretatie

De interpretatie van η (**eta**): wanneer eta 0,258 is, wil dat zeggen dat er een zwak verband is tussen de afhankelijke variabele en de onafhankelijke variabele op nominaal niveau (bijvoorbeeld sekse).

De interpretatie van η^2 (**eta-kwadraat**) is gelijk aan de interpretatie van R^2 . Wanneer η^2 0,056 is, wil dat zeggen dat de onafhankelijke variabele sekse 5,6% van de varia(n)tie verklaard in de afhankelijke variabele. Dit is niet heel veel, aangezien 94,4% onverklaard blijft.

Berekening

Wanneer je de sterkte van een verband tussen bijvoorbeeld geslacht en een andere interval of ratio variabele wilt weten, gebruik je eta (η). Eta is de wortel uit eta-kwadraat:

$$\eta = \sqrt{\eta^2}$$

Eta-kwadraat wordt niet alleen op dezelfde wijze als de proportie verklaarde variantie (R^2) geïnterpreteerd, hij wordt ook op dezelfde wijze berekend. Zie hiervoor het vorige hoofdstuk.

De formule voor η^2 :

$$\eta^2 = \frac{E_1 - E_2}{E_1}$$

JoHo Samenvatting – Beschrijvende Statistiek

E_1 is ook hier het aantal voorspellingsfouten dat gemaakt wordt, wanneer je voorspellingen doet op basis van de informatie van y . De afhankelijke variabele y is op interval- of op rationiveau gemeten, waardoor je het rekenkundig gemiddelde kunt gebruiken van y (\bar{y}).

De formule voor E_1 is:

$$E_1 = \sum (y - \bar{y})^2$$

E_2 berekent het aantal voorspellingsfouten dat je maakt wanneer je ook de informatie van de nominale variabele x gebruikt. De waarde van y kan worden voorspeld door voor elke x -waarde apart \bar{y} uit te rekenen. Dat gemiddelde gebruik je vervolgens als voorspeller voor die speciale waarde van x . Het aantal gemaakte fouten voor die speciale waarde van x is vervolgens weer de kwadratensom (\sum) van alle afwijkingen van dat gemiddelde. Voor elke waarde van x moet dit herhaald worden. Alle kwadratensommen die hier uitkomen, moeten bij elkaar opgeteld worden om E_2 te krijgen ($\sum\sum$).

De formule voor E_2 :

$$E_2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{ij})^2$$

Berekening in SPSS

Zoals je ziet is het erg ingewikkeld en veel werk om eta met de hand uit te rekenen. SPSS kan het in ieder geval een stuk sneller. Eerst ga je naar Analyze → Compare Means → Means. In het Means venster geef je aan wat de onafhankelijke en de afhankelijke variabelen zijn. In Options kan aangegeven worden welke informatie je wilt weten en kan onderaan eta aangevinkt worden (Anova table and eta).

Uitleg output

In de output van SPSS zijn E_1 en E_2 terug te vinden in de ANOVA Table. In de eerste kolom vinden we onder Sum of Squares de totale variatie (SStotal), dit is E_1 . Hier direct boven vinden we de onverklaarde variatie bij Sum of Squares Within Groups (SSwithin), dit is E_2 . Bovenaan de eerste kolom wordt eveneens $E_1 - E_2$ gegeven, namelijk Sum of Squares Between Groups (SSbetween). Dit is de verklaarde variatie.

Eta-kwadraat en eta kunnen met behulp van deze gegevens ook berekend worden met de eerder gegeven formule:

$$\eta^2 = \frac{E_1 - E_2}{E_1} = \frac{(SS_{\text{total}} - SS_{\text{within}})}{SS_{\text{total}}}$$

JoHo Samenvatting – Beschrijvende Statistiek

In de tabel Measures of Association, onder de ANOVA tabel, worden η en η^2 ook gewoon gegeven.

Associatiemaat kiezen

De keuze voor een associatiemaat is grotendeels afhankelijk van de meetniveaus van de variabelen die gebruikt worden. Wanneer één van de twee variabelen in je kruistabel nominaal is, kies je normaal gesproken voor een associatiemaat op hetzelfde niveau.

Nominaal niveau	Ordinaal niveau	Interval- en rationiveau
Cramers V	gamma	correlatiecoëfficiënt
phi	Somers' d	proportie verklaarde variantie
Goodman en Kruskals tau	Kendalls tau-b	gestandaardiseerde regressiecoëfficiënt
lambda	Spearman's rho	

Vervolgens ga je kijken of de relatie tussen de variabelen symmetrisch of asymmetrisch is. Kan je op basis van je theoretisch kader stellen dat een variabele invloed heeft op een andere? Dan is de relatie asymmetrisch en kunnen een onafhankelijke en een afhankelijke variabele aangewezen worden.

In speciale gevallen kan gebruik gemaakt worden van eta en eta-kwadraat, namelijk als de onafhankelijke variabele nominaal is en de afhankelijke variabele minimaal intervalniveau heeft.

Het kiezen van de juiste associatiemaat is daarnaast afhankelijk van de probleemstelling die je geformuleerd hebt. Wat wil je precies weten en op welke onderzoeksvraag wil je een antwoord vinden? Dit kan zelfs een reden zijn om een associatiemaat te gebruiken die je op basis van de (a)symmetrie of het meetniveau van de variabelen niet zou kiezen.

Bovendien kunnen de kenmerken van associatiematen van belang zijn bij het kiezen van associatiematen. Zo zijn de r en de β gebaseerd op het rekenkundig gemiddelde en Spearman's rho op de rangordening van waarden. Daarnaast is Kendall's tau-b bijna altijd een betere keuze dan gamma, omdat die maat specifiek is. Er wordt bij de tau-b immers rekening gehouden met geknoopte paren. Ditzelfde onderscheid kan gemaakt worden tussen de lambda en de Goodman en Kruskals tau. Voor de Goodman en Kruskals tau gebruik je meer gegevens dan in de lambda, waardoor die maat specifiek is en dus vaak meer geschikt is dan de lambda.

Cramers V is gebaseerd op de geobserveerde en verwachte waarden in een kruistabel. Op het moment dat in veel cellen in de kruistabel geen waarde is aangegeven, is het goed om de variabelen te hercoderen naar bijvoorbeeld weinig (in plaats van antwoordmogelijkheid 1 tot en met 5) en veel (in plaats van antwoordmogelijkheid 6 tot en met 10). Door het hercoderen kunnen van alle cellen geobserveerde waarden gemaakt worden.