
Hoorcollege 4

Correlaties

Het herkennen van het meetniveau en de variabelen is erg belangrijk. In een correlatiestudie wordt gekeken of er een relatie is tussen twee variabelen. Bijvoorbeeld of er een relatie is tussen het aantal uren dat je voor het tentamen leert en je tentamen cijfer? Of is er een relatie tussen het aantal Facebook-vrienden en de studiecijfers? Dit kan worden beantwoord aan de hand van een correlatiestudie.

Deze studies kijken naar lineaire relaties tussen twee variabelen. Deze relatie wordt beschreven aan de hand van een correlatiecoëfficiënt voor de sterkte, richting en significantie van de correlatie. Om te beginnen worden er twee verschillende correlaties uitgelegd. Wanneer de twee onderzochte variabelen van interval/ratio meetniveau zijn wordt dit de Pearson correlatie genoemd. Wanneer twee onderzochte variabelen van ordinaal meetniveau zijn wordt dit de Spearman correlatie genoemd. Er wordt alleen gekeken of er überhaupt een relatie is. Er wordt dus niet gekeken of de ene variabele de andere variabele beïnvloedt of veroorzaakt. Correlatie is dus geen causatie.

Als we naar een correlatie gaan kijken moet er eerst gekeken worden naar een grafiek. Er moet namelijk sprake zijn van een lineaire trend. Deze grafiek wordt een spreidingsdiagram genoemd. De relatie tussen twee variabelen kan ook een ander patroon vormen bijvoorbeeld een kromme. Er kan dan geen correlatie zijn.

Er moet vervolgens ook gekeken worden naar de richting van de relatie. Het gaat hierbij om een algemene trend. Er kan sprake zijn van een positieve en negatieve richting. Een positieve richting betekent dat als de X-variabele toeneemt de Y-variabele ook toeneemt. De twee variabelen bewegen dus in dezelfde richting. Een negatieve relatie geeft een toename in de X-variabele en een afname in de Y-variabele. De 2 variabelen bewegen in tegengestelde richting.

Pearson correlatiecoëfficiënt

De Pearson correlatie meet de sterkte en de richting van het lineaire verband tussen twee variabelen op het interval/ratio meetniveau. Het gaat dus alleen over de lineaire relatie. De correlatie wordt aangeduid met een r . r neemt waardes aan tussen de -1 en $+1$. Het teken van de correlatie geeft de richting aan. De waarde van de correlatie geeft de sterkte aan:

- Waarden die dicht bij 0 liggen geven een zwakke relatie aan.
- Waarden die dicht bij $+1$ liggen geven een sterk verband aan.
- Een waarde van $+1$ geeft een perfect lineair verband aan.

De Pearson correlatie gaat meten hoeveel van de spreiding in X en de spreiding van Y een gezamenlijke spreiding is. De gezamenlijke spreiding van X en Y delen door de spreiding van X en Y apart geeft de correlatiecoëfficiënt.

In symbolen ziet de formule er als volgt uit:

$$r = \frac{SP}{\sqrt{SSx \cdot SSy}}$$

De spreiding van X en Y wordt gemeten met de kwadratensom (SS). Voor de gezamenlijke spreiding gebruiken we niet een kwadratensom maar een productensom (SP=som of products).

De formule om SP te berekenen ziet er als volgt uit:

$SP = \sum(X - M_x)(Y - M_y)$. De correlatie in de populatie wordt aangegeven door de Griekse letter rho ofwel ρ .

De berekeningen van SS_x en SP op slide 20 komen apart op BlackBoard te staan.

SPSS

Je geeft aan welke variabelen je gebruikt om de correlatie weer te geven. SPSS gaat alle correlaties uitrekenen waardoor er een correlatiematrix ontstaat op slide 22. De correlatie X en X is altijd 1. De correlatie tussen Y en Y is altijd 1. Dit komt omdat het perfect lineair is. De correlatie tussen X en Y en Y en X is ook hetzelfde. De p waarde in de tabel is om te kunnen toetsen.

Stappen voor het toetsen

Allereerst worden de hypothesen bepaald en het significante niveau gekozen. De hypothese wordt gemaakt met het symbool rho.

$H_0: \rho = 0$ of $H_1: \rho \neq 0$.

De tweede stap zijn de kritieke waarden berekenen. Als H_0 waar is verwachten we dat de steekproefcorrelatie r dicht bij 0 zal vallen. Dan zullen we H_0 niet verwerpen als de r klein is en we zullen H_0 verwerpen als de r groot is. We zullen H_0 dan verwerpen als de r groter of gelijk is aan de r -kritiek. De kritieke waarden staan in tabel bij B.6. De beslissingsregel bij deze toets is dat we H_0 verwerpen als $r \geq r_{\text{kritiek}}$. Hierbij worden de vrijheidsgraden van $n - 2$ gebruikt.

Meestal wordt bij deze correlaties gebruik gemaakt van een tweezijdige toets.

Bij de derde stap wordt de toetsingsgrootte uitgerekend. De toetsingsgrootte hoeft bij de correlatie niet opnieuw uitgerekend te worden, want deze is gelijk aan de correlatiewaarde.

De vierde stap is de beslissing nemen over H_0 , waarbij de H_0 wel of niet verworpen wordt. Bij het eerste voorbeeld uit de PowerPoint mag de nulhypothese niet verworpen worden. De laatste stap is het rapporteren van de conclusie. Er moet op drie decimalen na nauwkeurig gerapporteerd worden. Volgorde van de rapportage is als volgt; allereerst de r . Hierbij moet een + of - vermeld worden. Daarna volgt de n en daarna de p . Tot slot of het een eenzijdige of tweezijdige toetsing is. Op slide 28 is een voorbeeld te zien hoe de uitkomsten gerapporteerd moeten worden volgens APA.

Let op, de factoren met een grote invloed op de correlatiecoëfficiënt, zoals bijvoorbeeld uitschieters of een beperkt bereik (deel van de scores die je bekijkt, subgroepen in plaats van het totaal) van data in de steekproef. Daarom is het erg belangrijk om eerst een grafiek te maken.

Spearman correlatiecoëfficiënt

De Spearman correlatiecoëfficiënt meet de sterkte en de richting tussen 2 ordinale variabelen. Ordinale variabelen geven alleen een locatie zoals bijv. rang aan: slechtste, beter en beste. Deze twee variabelen kunnen oorspronkelijk al ordinaal gemeten zijn; of ordinaal gemaakt zijn door middel van rangscores om een non-lineaire relatie recht te trekken. Dit kan alleen maar als de kromme alleen maar stijgend is of alleen maar dalend is. De Spearman correlatiecoëfficiënt wordt berekend met dezelfde formule als de Pearson correlatie. Alleen wordt deze berekend aan de hand van rangscores (de scores op volgorde zetten en een rang toedelen). Een voorbeeld hieronder. Eerst onder X staan de getallen op een random volgorde en daarnaast worden ze op basis van klein naar groot opgedeeld in rangscores.

| X | Rangscores |
|----|------------|
| 2 | 1 |
| 6 | 3 |
| 11 | 5 |
| 9 | 4 |
| 16 | 6 |
| 4 | 2 |

Een non-lineaire (kromme) relatie kun je rechtmaken door rangscores te berekenen. Als een lineaire lijn niet helemaal de goede relatie weergeeft, dan kunnen er geen rangscores berekend worden. Bij de kromme wel, omdat er een consistente toename is.

Als er dan sprake is van een knoop (tie) wat betekent dat er twee dezelfde X-en zijn, krijgen deze geen verschillende rangscore maar beiden de helft van de rangscore. Na het opnieuw maken van een grafiek is zichtbaar dat de kromme recht getrokken is.

De Spearman correlatiecoëfficiënt wordt berekend met dezelfde formule als de Pearson, aan de hand van de rangscores voor X en Y. Ook voor de Spearman correlatie kunnen we een toets doen. Deze werkt hetzelfde als de vorige correlatietoets. Hiervoor moet echter wel een andere tabel gebruikt worden, namelijk B.7. Let op, een verschil met B.6 is dat de kritieke waarden gebaseerd zijn op n en niet op $df = n - 2$. Ook Spearman's rho kun je in de SPSS-output vinden. Hoe dit gerapporteerd moet worden staat aangegeven op slide 49.

Andere correlaties

Er zijn nog twee andere correlaties, deze gebruiken dichotome variabelen (variabelen met maar twee waarden en meestal is dat 0 en 1). Alles kan dus in een ja nee variabelen veranderd worden. Er kan bijvoorbeeld gekeken worden naar sekse, studie (wel of niet in Utrecht). Beide correlaties geven antwoord op de volgende vraag: komen de hoge (of lage) waarden op de ene variabele overeen met hoge (of lage) waarden op de andere variabele? Als het ene negatief is moet het andere dat dus ook zijn en andersom. De phi-coefficient correlatie meet het verband tussen twee dichotome variabelen. Formule blijft hetzelfde alleen een andere tekenen. Let op! Zowel X en Y bestaan alleen uit 0 en 1.

De punt-biseriële correlatie meet de sterkte van het verband tussen een variabele op interval of ratio niveau en wanneer het een dichotome variabele is. De formule blijft altijd hetzelfde.

De r wordt dan r_{PB} . De Y bestaat alleen uit 0 of 1.