

10. Moderatie, mediatie en nog meer regressie

Voordat je moderatie en mediatie analyses gaat uitvoeren in SPSS, kun je het best een extra dialog box installeren, *PROCESS*. Volg hiervoor de stappen op pagina 393 en 394.

Moderatie: interacties in regressie

Het conceptuele model

Een gecombineerd effect van twee of meer predictorvariabelen heet als concept moderatie, en in statistische termen een interactie-effect. Een moderator beïnvloedt de relatie tussen een predictor en de uitkomst.

Een voorbeeld is de relatie tussen het spelen van videogames en agressie. Hardvochtigheid kan een moderator zijn. Hierbij wordt hardvochtigheid gemeten als een dichotome variabele, wel of niet hardvochtig. Een moderatoreffect kan dan als volgt zijn: Alleen voor hardvochtige mensen heeft het spelen van videogames invloed op agressie. Voor niet-hardvochtige mensen is er geen verband.

Het verschil tussen wel of geen effect is de simpelste vorm van moderatie. Je kunt ook andere vormen van verandering hebben door de moderator. Denk bijvoorbeeld aan een versterking van het effect of een verandering van richting.

Als de moderatorvariabele continu is, kun je kijken hoe de relatie tussen de predictor en uitkomst verandert als een functie van de moderator. In het voorbeeld vergelijk je dan de helling van het regressievlak kijkend naar hoge of lage mate van hardvochtigheid. Als er geen moderatie-effect is, vertoont het regressievlak een gelijke helling bij hoge en lage mate van hardvochtigheid. Als er wel een moderatie-effect is, is de helling verschillend bij een verschillende mate van hardvochtigheid.

Het statistische model

Statistisch gezien voorspellen we de uitkomst vanuit de predictorvariabele, de moderator en de interactie tussen de predictor en de moderator. Om deze interactie-term te kunnen gebruiken, moeten de predictor en moderator ook in het model zitten. In de trant van het standaard model in de statistiek, schrijf je een moderatie zo op:

$$\text{Uitkomst} = (b_0 + b_1 \text{predictor} + b_2 \text{moderator} + b_3 \text{interactie}) + \text{error}$$

De b-waardes hebben in dit model een specifieke betekenis, het is de regressie op de uitkomst van die predictor als de andere predictor nul is. In het huidige voorbeeld kan dat, je kunt nul uren besteden aan het spelen van videogames, en je kunt ook een score van nul hebben op de hardvochtigheidsschaal. Echter, als je niet hardvochtigheid meet, maar hartritme, dan is een score van nul op z'n minst zorgwekkend. Daarom worden predictoren vaak getransformeerd met *grand mean centring*, centreren rond het overkoepelende gemiddelde.

Centreren betekent dat je een variabele transformeert naar deviaties rond een bepaald punt. Het punt dat doorgaans gekozen wordt, is het overkoepelende gemiddelde (over de condities heen). Dus je krijgt het gecentreerde punt door de score af te trekken van het overkoepelende gemiddelde.

Het centreren heeft geen invloed op de b-waarde voor de 'highest-order' predictor, maar wel voor de 'lower-order' predictors.

De higher order predictor is het effect met de meeste variabelen, dus in dit geval de interactieterm (gamen x hardvochtigheid). Dus b_3 verandert niet, maar b_1 en b_2 wel. Nu staat de b-waarde voor het effect van die predictor als de andere predictor de gemiddelde waarde heeft.

Wiskundig gezien is een interactie gewoon een vermenigvuldiging van de effecten van de predictoren. Daarom wordt het ook genoteerd als variabele 1 x variabele 2. Je kunt dit dus zelf als variabele maken.

Moderatie is een significante interactie tussen de twee variabelen. Je weet alleen daarmee nog niet precies wat het effect van de moderator is. daarvoor voer je een *eenvoudige richtingscoëfficiënten analyse (simple slopes analysis)* uit. Deze analyse berekent regressievergelijkingen van de predictor op de uitkomstvariabele bij hoge, gemiddelde en lage scores van de moderator. Je vergelijkt deze b-waardes op significantie en richting.

Je kunt ook de relatie tussen de predictor en de uitkomst bekijken op heel veel waardes van de moderator. Hiermee krijg je een ‘zone van significantie’, wat bestaat uit twee waardes van de moderator. Tussen deze twee waardes is het verband tussen predictor en uitkomstvariabele niet significant, onder en boven die waardes wel.

SPSS

Aangezien moderatie gewoon een regressie-analyse is, kun je het in principe gewoon uitvoeren via het standaard scherm. Je moet eerst de predictor en moderator centreren, de interactieterm creëren, en dan een forced entry regressie uitvoeren met de gecentreerde predictor, gecentreerde moderator en de interactie tussen die twee als voorspellers.

Met het *PROCESS* hulpmiddel hoef je echter niet zelf te centreren, hoef je zelf de interactieterm niet te creëren en voert het een simple slopes analyse uit. In dit scherm kun je de onafhankelijke variabele, de uitkomstvariabele en moderators invullen. De moderators vul je in bij M variable(s). Bij model number moet je 1 selecteren, dat staat voor moderatie. Standaard staat hij op 4 (mediatie), dus niet vergeten te veranderen!

Bij Options kun je kiezen om de predictor en moderator te centreren (Mean center for products). De optie heteroscedasticity-consistent SEs zorgt ervoor dat je niet hoeft na te denken over of er heteroscedasticiteit is in het model. OSL/ML confidence interval geeft je een betrouwbaarheidsinterval. Bij Conditioning kun je kiezen om de simple slopes analyse uit te voeren met het gemiddelde en 1 standaardafwijking boven en onder het gemiddelde, of met percentielen. Als je Johnson-Neyman selecteert, krijg je een zone van significantie. Generate data for plotting betekent dat je een grafiek krijgt.

De output verschijnt als tekst in plaats van tabellen. Zie bladzijde 404 en 405. Er worden drie verschillende regressieanalyses getoond, namelijk bij een hoge, gemiddelde en lage waarde van de moderator (in output 10.2). Het effect is de b-waarde. In output 10.3 is de zone van significantie te zien. Tenslotte is er een grafiek van de simple slopes analyse, waarbij drie lijnen worden weergegeven, een van lage waardes van de moderator, een van hoge waardes en een van de gemiddelde waarde van de moderator. Dit maakt het makkelijker te interpreteren.

Mediatie

Je spreekt van *mediatie* als de relatie tussen een predictorvariabele en de uitkomst (c) verklaard wordt door de relatie met een derde variabele, de *mediator*. De predictor voorspelt dan de mediator, en de mediator voorspelt de uitkomst.

De relatie tussen predictor en uitkomst is waarschijnlijk anders met de mediator in het model, dus die noem je c' . De letters, zoals te zien in de figuren op pagina 408 en 409, zijn regressiecoëfficiënten.

Er is sprake van mediatie als de relatie tussen predictor en uitkomstvariabele minder sterk wordt door het toevoegen van de mediator. De regressieparameter voor c' is dan kleiner dan voor c . Als $c' = 0$ is, is er perfecte mediatie, bij toevoegen van de mediator bestaat er geen verdere relatie meer tussen predictor en uitkomst.

Bij mediatie heb je een *direct effect*, de relatie tussen de predictor en de uitkomst, terwijl je controleert voor de mediator, en je hebt een *indirect effect*, het effect van de predictor op de uitkomst via de mediator.

Het statistische model

Het statistische model bij mediatie is hetzelfde als het conceptuele model. Mediatie is gebaseerd op drie regressiemodellen:

- Een regressie die de uitkomst voorspelt vanuit de predictor. De regressiecoëfficiënt geeft ons de waarde van c .
- Een regressie die de mediator voorspelt vanuit de predictor. Dit geeft a in het schema op pagina 408.
- Een regressie die de uitkomst voorspelt vanuit zowel de predictor als de mediator. De regressiecoëfficiënt voor de predictor is c' , de regressiecoëfficiënt voor de mediator is b .

Deze modellen testen de vier condities van mediatie, namelijk:

- De predictorvariabele moet de uitkomst significant voorspellen.
- De predictor moet de mediator significant voorspellen.
- De mediator moet de uitkomst significant voorspellen.
- De predictor moet de uitkomstvariabele minder sterk voorspellen in het derde model dan in het eerste model.

Dit laatste criterium zorgt voor onduidelijkheid. Perfecte mediatie komt nauwelijks voor, meestal is er alleen een minder sterke relatie tussen predictor en uitkomst door de mediatie. Maar het is daarmee onduidelijk met hoeveel deze relatie moet afnemen om te spreken van mediatie.

In de praktijk zijn er veel mensen die kijken naar een verschil in significantie, wel een significante relatie in model 1, maar niet in model 3. Dat is alleen niet een goede aanpak, omdat een klein verschil in regressieparameters toch net een verschil in significantie kan maken en een groot verschil in regressieparameters geen verschil in significantie kan maken.

Een alternatieve methode is het schatten van het indirecte effect, of dat significant is. Dat kan met de *Sobel test*. Het is echter beter om betrouwbaarheidsintervallen voor het indirecte effect te maken met de bootstrap methode.

Effectgrootte

De makkelijkste manier om de effectgrootte te berekenen, is door te kijken naar het gecombineerde effect van a en b , de regressiecoëfficiënten van de relatie tussen predictor en mediator / mediator en uitkomst. Dit doe je door a en b te vermenigvuldigen.

Hiermee krijg je een niet-gestandaardiseerde regressiecoëfficiënt. Om het indirecte effect gedeeltelijk te standaardiseren, deel je het gecombineerde effect ab door de standaardafwijking van de uitkomst.

Om het indirecte effect helemaal te standaardiseren, moet je ook de standaardafwijking van de predictor erbij nemen:

$$\text{Indirect effect (gestandaardiseerd)} = \frac{ab}{S_{\text{uitkomst}}} \times \text{Spredictor}$$

Dit wordt soms ook de *index van mediatie* genoemd. Deze meting kan vergeleken worden met andere mediatie modellen.

Je kunt de effectgrootte ook berekenen door te kijken naar de grootte van het indirecte effect vergeleken met het totale effect van de predictor of het directe effect. Je kunt bijvoorbeeld het indirecte effect (ab) delen door het totale effect (c). Je kunt ab ook delen door het directe effect (c'). Het is echter beter om dit niet te gebruiken, aangezien het lijkt op een proportie, maar dat niet is, en omdat het alleen in heel grote steekproeven bruikbaar is.

Voor mediatie kan ook R^2 worden berekend. Hierbij hoeft de meting echter niet tussen de 0 en de 1 te vallen. κ^2 (kappa-squared) is wel een proportie, en kun je dus zo interpreteren. Dit vergelijkt het indirecte effect met het maximale indirecte effect dat je had kunnen krijgen, gezien het design van je onderzoek.

Met *PROCESS* krijg je al deze effectgroottes. Je kunt het beste alleen het ongestandaardiseerde en gestandaardiseerde indirecte effect gebruiken, en κ^2 .

SPSS

Via Analyze – Regression –PROCESS kom je bij het juiste scherm. De mediator voer je in bij M Variable(s). Voor mediatie moet je model 4 aanklikken bij Model number (die is standaard al geselecteerd). Een bootstrap voor indirecte effecten is handig, de standaard opties zijn prima, maar je kunt ook om percentielen vragen.

Bij Options kun je de laatste vier opties aanvinken. De laatste optie, Compare indirect effect, is alleen nuttig als je meerdere mediators hebt, want het vergelijkt hoe relatief belangrijk de mediators zijn in het verklaren van de relatie tussen predictor en uitkomst. Als je slechts één mediator hebt, maakt deze optie geen verschil in output.

De Conditions knop hoeft je niet te gebruiken bij mediatie.

Ook hier verschijnt de output als tekst. Op pagina 415 zie je het eerste deel, waarbij je (afgekort) ziet welke variabelen je gebruikt, en daaronder de samenvatting van het model. Als je de optie Total effect model hebt aangeklikt, krijg je een tabel met daarin het model zonder de mediator.

Output 10.7 op pagina 417 is het belangrijkste deel van de output, met de resultaten van het indirecte effect. Eerst wordt het totale en het directe effect opnieuw genoemd, daaronder staat het indirecte effect. BootLLCI en BootULCI zijn de grenzen van het bootstrap betrouwbaarheidsinterval.

De rest van output 10.7 krijg je alleen als je de optie Effect size hebt aangeklikt. De κ^2 kun je interpreteren als een percentage van het maximale effect. In het voorbeeld heeft het indirecte effect 4,1% van het maximum, wat een vrij klein effect is. Tenslotte staat er in output 10.8 op pagina 418 de output van de Sobel test.

Categorische predictors in regressie

In het vorige hoofdstuk kwam naar voren dat je een categorische predictor kunt gebruiken in een regressiemodel als er slechts twee categorieën zijn. De ene categorie codeer je dan als 0, de andere als 1. Maar als je een categorische predictor met meer categorieën hebt, moet je dit omzetten naar meerdere dichotome predictoren. Dat heet dummy coderen.

Dummy variabelen zijn meerdere variabelen die met alleen nullen en enen gecodeerd zijn.

Er zijn 8 basisstappen om dummy variabelen te creëren:

- Tel het aantal groepen dat je wil coderen, en doe dat aantal min 1.
- Maak net zoveel nieuwe variabelen als wat je geteld hebt in stap 1.
- Kies een groep als basisgroep, waarmee je de andere groepen vergelijkt. Dit is meestal de controlegroep of de groep die de meerderheid vertegenwoordigt.
- Zodra je de basisgroep hebt, ken je die groep de waarde 0 toe bij elke dummy variabele.
- Bij de eerste dummy variabele codeer je de eerste groep die je met de basisgroep wil vergelijken met een 1, en alle andere groepen een 0.
- Bij de tweede dummy variabele codeer je een 1 bij de tweede groep die je met de basisgroep wil vergelijken, en bij alle andere groepen een 0.
- Herhaal dit tot je het bij alle dummyvariabelen hebt gecodeerd.
- Stop alle dummy variabelen in de regressieanalyse, in hetzelfde block. Hiervoor moet je de Enter methode gebruiken.

Uiteindelijk heeft elke groep bij één dummyvariabele een 1 en bij alle andere variabelen een 0, behalve de basisvariabele, die heeft bij alle dummyvariabelen een 0.

SPSS

Voor SPSS ga je naar Transform - recode into different variables. De variabele die je wil coderen, zet je in het veld Numeric variable -> output variable. Bij het kopje Output variable kun je een nieuwe dummyvariabele maken. Klik op Change om het door te voeren. Bij Old en new values codeer je een bepaalde categorie die je wil vergelijken met de basisgroep als 1. De andere groepen codeer je met 0 (hiervoor kun je de optie All other values gebruiken). Klik op Add. De schermen staan ook op bladzijde 421 en 422.

De output begint met een samenvatting van het model en een ANOVA tabel. De tabel met Coefficients is het belangrijkste met dummyvariabelen. Hierin kun je per dummyvariabele vinden wat de coëfficiënten zijn en of er een significant verschil is tussen een bepaalde groep en de basisgroep.