

10. Inleiding in Meervoudige Relaties

Verbanden en causaliteit

In veel wetenschappelijke studies wordt gekeken naar meer dan één of twee variabelen. Multivariate methoden worden dan gebruikt. Veel studies gaan over causale relaties tussen variabelen. Maar het vaststellen van een causale relatie is niet gemakkelijk. Het schijnbaar causale verband kan immers ook worden veroorzaakt door een andere variabele. We maken daarom gebruik van statistische controle, waarbij gekeken wordt of een verband tussen variabelen verandert of zelfs verdwijnt wanneer er ook gecontroleerd wordt voor andere variabelen.

Een causaal verband bestaat uit een verklarende variabele (X) en een responsvariabele (Y), waarbij X de oorzaak is van Y. In schema: $X \rightarrow Y$. Om een causaal verband vast te stellen, moet dit voldoen aan drie criteria:

- Er moet een verband zijn tussen de twee variabelen.
- De verklarende variabele moet vooraf gaan aan de respons variabele.
- Andere verklaringen moeten zijn uitgesloten.

Controlevariabelen

Een belangrijk component is dat andere verklaringen moeten zijn uitgesloten. Dit is vaak heel lastig. Het uitsluiten van de invloed van andere variabelen op jouw causale verband doe je door voor die andere variabele te 'controleren'. Je noemt deze variabelen dan ook controle variabelen. Je houdt dan de waarde van de controle variabelen constant.

Bij een kwantitatieve en een categorische variabele beschrijf je het verband door de gemiddelden te vergelijken. Bij twee categorische variabelen maak je kruistabellen om het verband te onderzoeken.

Wanneer je een causaal verband vaststelt, en je hebt daarbij niet gecontroleerd voor een variabele die wel van invloed is, simpelweg omdat je er bijvoorbeeld niet van op de hoogte bent dat die variabele invloed heeft op het causale verband, is er sprake van een 'lurking' variabele. Een lurking variabele is niet gemeten, maar wel van invloed op het causale verband.

Typen van multivariate relaties

We weten dat associaties tussen twee variabelen heel erg kan veranderen, of zelfs verdwijnen wanneer we controleren voor andere variabelen. Hieronder een aantal typen van deze multivariate relaties.

Valse relaties (spurious associations)

Er is sprake van deze relatie wanneer zowel de verklarende als de responsvariabele afhankelijk is van een derde variabele (X2), en dat het verband tussen de verklarende (X1) en responsvariabele (Y) verdwijnt wanneer je controleert voor X2. Wanneer je dit tegenkomt, bestaat er dus geen causale relatie tussen X1 en Y. In schema:



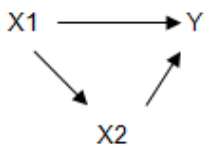
Ketting relaties (chain relationships)

Hierbij 'veroorzaakt' de verklarende variabele (X1) een derde variabele (X2), die op zijn beurt weer de responsvariabele (Y) 'veroorzaakt'. Deze derde variabele (X2) wordt ook wel de mediator genoemd. In schema: $X1 \longrightarrow X2 \longrightarrow Y$

Merk op dat bij bovenstaande relaties het verband tussen X1 en Y verdwijnt zodra je controleert voor X2. Het gebeurt alleen op andere manieren. Het verschil ligt in de causale volgorde. Bij een 'onechte' relatie gaat X2 vooraf aan zowel X1 als Y. Bij een ketting relatie bemiddelt X2 tussen X1 en Y.

Meerdere oorzaken

In de werkelijkheid hebben responsvariabelen bijna altijd meer dan één oorzaak. Soms zijn deze oorzaken onafhankelijk, maar meestal houden deze oorzaken ook weer verband met elkaar. Dit betekent dat bijvoorbeeld X1 behalve een direct effect op Y ook een indirect effect op Y kan hebben via X2. In schema:



Onderdrukker (suppressor variable)

Hierbij lijkt er geen verband te zijn tussen X1 en Y, totdat je controleert voor X2. X2 is hier een 'onderdrukkende variabele'. Dit kan gebeuren wanneer bijvoorbeeld X2 positief gecorreleerd is met Y, maar ook negatief gecorreleerd met X1. Daarom, zelfs wanneer er in eerste instantie geen verband lijkt te zijn tussen twee variabelen, is het verstandig om alsnog voor andere variabelen te controleren.

Statistische interactie

Er is sprake van statistische interactie tussen X1 en X2 en hun effect op Y, wanneer het echte effect van een van de predictoren op Y verandert, als de waarde van de andere predictor verandert. Bijvoorbeeld wanneer we kijken naar het inkomen van mannen en vrouwen, en de opleiding die zij hebben gehad. We weten dat mannen gemiddeld meer verdienen dan vrouwen. En we weten dat hoog opgeleiden meer verdienen dan laagopgeleiden. Er is sprake van statistische interactie wanneer opleiding leidt tot meer inkomen, meer voor mannen dan voor vrouwen. Bijvoorbeeld wanneer mannen per opleidingsjaar 5% in uurloon stijgen en vrouwen per opleidingsjaar maar 3% in uurloon stijgen. Er is dan interactie tussen opleiding en geslacht in het effect op uurloon. Later leer je hoe je kunt zien of dat interactie effect significant is.