

Hoofdstuk 10. De factoranalyse

Naast de toevallige fouten die gemaakt kunnen worden in een meting, zoals in het vorige hoofdstuk, kunnen ook systematische fouten gemaakt worden. Er is dan niet een probleem met de betrouwbaarheid van je meting, maar met de **validiteit**. Meet je wel wat je wilt meten?

Ook hier geldt dat makkelijke vragen in een enquête geen problemen zullen opleveren voor de validiteit van je studie. Onderzoek naar complexe verschijnselen of abstracte begrippen kan wel problemen opleveren, omdat de meting hiervan minder voor de hand ligt.

Manifeste en latente variabelen

Bij de operationalisatie van de concepten waarnaar je onderzoek doet, moet je goed nadenken over de juiste vragen of stellingen die je gebruikt in de enquête.

Als voorbeeld nemen we weer het concept 'hang naar sensatie' waarvoor 8 verschillende stellingen waren geformuleerd. Zo kan gevraagd worden naar verschillende aspecten van 'hang naar sensatie'. Voorbeelden hiervan waren: "Ik wil een keer bungeejumpen" en "Ik kijk graag spannende films". Deze verschillende aspecten worden manifeste uitingen genoemd van het concept 'hang naar sensatie'. Anders gezegd, deze 8 manifeste variabelen meten samen de latente variabele 'hang naar sensatie'. De voorwaarde om samen een latente variabele te vormen, is dat de correlaties tussen de 8 manifeste variabelen hoog zijn.

Factoranalyse

Aan de hand van een factoranalyse kun je iets zeggen over de validiteit van metingen. Als je uit de factoranalyse kan aflezen dat de manifeste variabelen unidimensioneel zijn en dus dezelfde latente variabele meten, kun je de manifeste variabelen samenvoegen tot één schaalvariabele voor die latente variabele. Deze nieuwe variabele wordt dan factor of component genoemd en zo kan een groot aantal variabelen worden gereduceerd naar een hanteerbaarder aantal. Deze datareductie maakt het onderzoek eenvoudiger.

In een correlatiematrix is soms snel te zien of alle manifeste variabelen hoog met elkaar correleren en dus samen een latente variabele vormen. De gemeenschappelijke correlaties van de groep variabelen wordt een factor genoemd. Vervolgens kan de correlatie tussen deze factor en de afzonderlijke manifeste variabelen worden berekend.

Factorlading

De correlatie tussen de factor en een manifeste variabele heet de factorlading. Uit de analyse blijkt hoe sterk de verschillende manifeste variabelen laden op een factor. De factorlading varieert daarbij tussen de -1 (perfect negatief) en de +1 (perfect positief). Een factorlading van 0 houdt in dat er helemaal geen samenhang is. Een manifeste variabele met een factorlading van kleiner dan 0,30 wordt doorgaans niet verder meegenomen in het onderzoek.

Constructvaliditeit

Uit de analyse moet blijken in hoeverre dezelfde latente variabele wordt gemeten door een aantal manifeste variabelen. Dit geeft de constructvaliditeit aan van je onderzoek. Met andere woorden: meet je wat je wilt meten?

JoHo Samenvatting – Beschrijvende Statistiek

Wanneer de manifeste variabelen niet sterk samen blijken te hangen, zullen ze niet hetzelfde concept (latente variabele) meten. Je onderzoek is dan niet valide. Het is ook mogelijk dat er sprake is van meerdere factoren. Dit wordt hieronder verder uitgelegd.

Berekening in SPSS

Het laten uitvoeren van een factoranalyse doe je via Analyze → Dimension Reduction → Factor. Onder Variables zet je alle (manifeste) variabelen die je mee wilt nemen in je analyse. In het voorbeeld zouden dat de 8 variabelen zijn waarmee 'hang naar sensatie' gemeten werd.

Interpretatie

Wanneer blijkt dat alle manifeste variabelen inderdaad hetzelfde concept meten (sterk samenhangen), dan is je meting valide. De manifeste variabelen kunnen samengevoegd worden tot een nieuwe variabele: een schaalvariabele. Hoe dit in zijn werk gaat, is te lezen aan het eind van het vorige hoofdstuk. Het samenvoegen van manifeste variabelen tot één schaalvariabele maakt je werk een stuk overzichtelijker en wordt ook wel datareductie genoemd.

Meerdere factoren

Een factoranalyse kan ook uitwijzen dat er meerdere latente variabelen ten grondslag liggen aan de gebruikte manifeste variabelen. Uit je SPSS output blijkt dan dat de manifeste variabelen op twee of zelfs meer factoren laden. Een manifeste variabele met een factorlading op factor 1 van minder dan 0,30, kan op een tweede factor bijvoorbeeld wel hoog scoren.

We pakken het voorbeeld 'hang naar sensatie' er weer bij. Om dit concept te meten worden 4 stellingen gegeven met betrekking tot extreme sporten (bijvoorbeeld: "Ik wil een keer bungeejumpen"). De andere 4 stellingen vragen naar een hang naar sensatie in mediagebruik (bijvoorbeeld: "Ik kijk graag spannende films"). De kans is groot dat de vier vragen over extreme sporten sterk met elkaar correleren, maar dat de samenhang met de andere vier stellingen gering is.

In de output van SPSS zal dan bijvoorbeeld blijken dat er twee factoren ten grondslag liggen aan de 8 manifeste variabelen, namelijk: 'hang naar sensatie op het gebied van sport' en 'hang naar sensatie op het gebied van mediagebruik'. De output zal in dat geval een Component Matrix geven met daarin de twee factoren (Components) en de bijbehorende factorladingen ofwel correlatie per manifeste variabele.

Explorerend en confirmatieve factoranalyse

Het is mogelijk dat een factoranalyse wordt uitgevoerd zonder dat onderzoekers van tevoren weten wat mogelijke latente variabelen zouden kunnen zijn. Als deze onderzoekers vervolgens op basis van de sterkst ladende manifeste variabelen een naam aan de gevonden factoren geven, is er sprake van een **explorerende factoranalyse**. Bij deze analyse gebruikt de onderzoeker de variabelen die op een bepaalde factor een hoge lading hebben om deze factor te interpreteren.

Als er echter al verwacht wordt op theoretische gronden dat met een aantal manifeste variabelen bepaalde latente verschijnselen te meten zijn, gaat het om een **confirmatieve factoranalyse**. Het gaat om een bevestiging van dat de vooraf benoemde latente variabelen met factoren in verband staan.

JoHo Samenvatting – Beschrijvende Statistiek

Factoren vinden

Aan de hand van verschillende factoren kunnen onderliggende factoren uit de onderlinge relaties tussen de manifeste variabelen kunnen worden geëxtraheerd. Ze worden als het ware uit elkaar getrokken en gesorteerd. Een voorbeeld van zo'n methode is de **principale componentenanalyse**. Hierbij spelen componenten een grote rol. Dit zijn de onderliggende dimensies die uit de samenhang tussen de variabelen worden geëxtraheerd.

Aan de hand van een spreidingsdiagram worden er altijd hetzelfde aantal componenten gevonden als manifeste variabelen en deze verklaren samen de gehele variantie. Er zijn drie soorten variantie per variabele. De specifieke variantie is de variantie die alleen door die specifieke variabele gemeten wordt en wat dat betreft dus uniek is voor die variabele. Daarnaast is de foutenvariantie een willekeurige variantie waarbij geen verband is met een andere bron van variantie, door bijvoorbeeld meetfouten of het gokken van antwoorden. Tot slot is de gemeenschappelijke variantie de variantie die de variabele gemeen heeft met andere variabelen en dus belangrijk is voor het vinden van de latente variabele. Bij een factoranalyse wordt aan de hand van een spreidingsdiagram gezocht naar een lijn ofwel factor, die zoveel mogelijk van deze gemeenschappelijke variantie verheldert.

Aantal factoren beperken

Aangezien SPSS alle variantie tussen de 8 variabelen wilt verklaren, zal het in de output mogelijk ook aangeven dat er 8 factoren zijn. Hier heb je echter weinig aan, want je wilde juist je data reduceren door variabelen met een gemeenschappelijke dimensie samen te voegen. In de output vindt je echter ook de eigenwaarde van iedere factor onder Total Initial Eigenvalues. In de kolom daarnaast staat hoeveel variantie iedere factor verklaart. De regel geldt dat alleen factoren worden meegenomen in de rest van de analyse met een eigenwaarde die hoger is dan 1. Dit wordt het criterium van Kaiser genoemd. De andere factoren laat je afvallen, waardoor je wel wat informatieverlies moet accepteren.

Een andere manier om te bepalen welke factoren je meeneemt en welke je laat afvallen, is het laten tekenen van een scree plot door SPSS. In de grafiek die je vervolgens krijgt, zie je een duidelijke knik in de lijn, deze wordt ook wel de 'elleboog' genoemd. De factoren voor deze knik neem je mee in je onderzoek en de factoren die na de knik liggen, vallen af.

Berekening in SPSS

Je volgt dezelfde stappen in SPSS zoals hierboven beschreven. Vervolgens klik je op Extraction en vink je Scree plot aan. In je scree plot zie je vervolgens de 'knik' in de grafiek. Aan de hand daarvan kan je bepalen hoeveel factoren je meeneemt in de rest van je onderzoek.

Rotatie

Soms laadt een manifeste variabele op twee factoren redelijk sterk. Je kan de variabele niet bij beide componenten indelen, maar hoe kies je? Dit doe je met behulp van een rotatie in SPSS. Dit houdt in dat SPSS op zoek gaat naar de optimale factorladingen voor beide factoren. Hij probeert de correlaties te maximaliseren op de factor waar zij sterk op laden en te minimaliseren op de factor waar zij zwak op laden.

De geroteerde componenten (factoren) die uit de matrix volgt, laat duidelijk zien dat de manifeste variabelen op één van de twee factoren sterk laden en op de andere zwak tot helemaal niet.

JoHo Samenvatting – Beschrijvende Statistiek

Berekening in SPSS

Om de manifeste variabelen optimaal te laten laden op de twee (of meer) factoren die je wilt berekenen, moet je SPSS een rotatie laten uitvoeren. Dit doe je door bij Rotation aan te geven dat je een Varimax rotatie wilt.