

STAPPENPLAN FACTORANALYSE

Factoranalyse is een statistische techniek die erop gericht is uit te maken hoeveel en welke dimensies er schuilgaan in een reeks gemeten indicatoren. Factoranalyse is in eerste instantie een explorerende techniek, die verbazingwekkend effectief is in het vinden van groepen samenhangende variabelen in je gegevens. Dit heet explorerende factoranalyse (EFA). Maar factoranalyse is ook geschikt ingeval je vooraf al een groepering van indicatoren veronderstelt. Je kunt het dan gebruiken om te toetsen of je veronderstelde hoeveelheid en aard van de dimensies inderdaad in de gegevens terug te vinden zijn. Dit heet confirmatieve factoranalyse (CFA) en is een belangrijke vorm van validiteitsanalyse.

- Twee hoofdvormen zijn (latente) factoranalyse (/extraction=paf) componentenanalyse (/extraction=pc). Hoewel deze vormen nogal verschillen in statistisch modellering, hebben ze vaak soortgelijke uitkomsten. Veel toepassers maken / kennen het verschil niet. SPSS heeft componentenanalyse als standaard en gebruikt het woord *component* inwisselbaar met *factor*.
- Factoranalyse is verbonden met een groot aantal termen die je elders in de statistiek niet of nauwelijks tegenkomt: eigenwaarde, rotatie, ladingen, communaliteit en nog zo wat. Vaak hebben deze termen wel parallellen met meer gangbare termen in de statistiek van correlatie- en regressie-analyse. Zie bijlage.
- Factoranalyse berust op de correlatiematrix tussen indicatoren. Je kunt de techniek daarom ook toepassen op een gegevens met partieel missende waarden (*pairwise deletion of missing values*).
- Componentenanalyse is ook geschikt voor, zelfs gericht op, het berekenen van een index als een (optimaal) gewogen gemiddelde van indicatoren. Dit heeft echter weinig voordelen boven een ongewogen gemiddelde en een groot nadeel: het kan slecht met missing values omgaan.
- Het is eigenlijk heel vreemd dat SPSS bij factoranalyse geen inferentiële statistiek (standard errors, F-toetsen, t-toetsen) levert. Voor dit type informatie kun je wel terecht in andere statistische pakketten (zoals LISREL of Stata).

Stap 0: Inspecteren en verder voorbereiden van de indicatoren.

- Bekijk via frequencies of de betrokken indicatoren ten minste ordinaal meetniveau hebben, missing values goed gedeclareerd zijn en er geen al te gekke uitschieters in zitten.
- Anders dan bij betrouwbaarheidsanalyse is het voor factoranalyse geen vereiste om indicatoren consistent te polen of ze te standaardiseren. Voor correcte interpretatie is het wel nodig dat je de poling van de indicatoren gemakkelijk correct kunt interpreteren. Ompolen kan helpen als je indicatoren in wisselende richting gevraagd zijn.
- Als je een confirmatoire vraagstelling hebt, is het verstandig de volgorde van de variabelen in de analyse aan je veronderstelde groepering van de indicatoren aan te passen.
- Vergelijk ingeval van missing values altijd een pairwise oplossing met een listwise oplossing. Bekijk vooral of je veel cases verliest door listwise deletion en je mogelijk een heel kleine dataset aan het analyseren bent.

Stap 1: Bepaal het aantal dimensies

- De vraag naar de hoeveelheid latente dimensies is typisch voor explorerende factoranalyse (EFA), maar ook als je een confirmerende vraagstelling (CFA) hebt, is er niets tegen om eerst een explorerende stap te doen.
- SPSS besluit op basis van het aantal eigenwaarden > 1.00 hoeveel dimensies nodig zijn om de indicatoren samen te vatten (het *kaiser-criterium*). Zeker als je een confirmerende vraagstelling hebt, is er niets tegen om meer of minder dimensies te bekijken. Dat doe je via `/criteria=factors(N)`.
- Een veel gebruikt criterium wordt ontleend aan de ‘scree-plot’: het aantal benodigde dimensies is knik-1. De screeplot is een grafiekje van de omvang van de opeenvolgende eigenwaarden: vanaf de knik bestaan de opeenvolgende dimensies uit ‘ruis’.
- Een handzame werkwijze kan zijn om achtereenvolgende analyses met verschillende hoeveelheid factoren te vergelijken en proberen te begrijpen hoe en waarom de (geroteerde) oplossingen van elkaar verschillen.

Stap 2: Kies een oblimin (=scheve) rotatie

- “Rotatie” is de stap waarmee je binnen een eenmaal gekozen hoeveelheid dimensies een zo eenvoudig mogelijke patroon van relaties tussen indicatoren en dimensie verkrijgt. Deze hoofdkeuze is die tussen Varimax (‘rechte’) en Oblimin (‘scheve’) rotatie. Bij scheve rotatie sta je toe dat de dimensies gecorreleerd zijn – en dat bijna altijd een heel natuurlijke veronderstelling. Als je namelijk werkelijk zou willen veronderstellen dat je dimensies ongecorrleerd zijn, is er ook geen reden om (multiple) factoranalyse te doen.
- Scheve rotatie zorgt wel voor ingewikkelde output dan rechte rotatie. Je moet de *pattern* matrix interpreteren, NIET de *structure* matrix en bovendien de correlaties tussen dimensies tegelijkertijd bekijken.

Stap 4: Bekijk en interpreteer de factorladingen.

- SPSS kent twee opties om de factorladingen overzichtelijker te maken. Met `/FORMAT=Blank(0.25)` kun je ladingen lager dan 0.25 onzichtbaar maken. Met `/SORT` kun je de indicatoren sorteren op grootte van de ladingen, hetgeen vooral bij EFA een nuttige stap kan zijn.
- Bij wetenschappelijke verslaggeving worden de hoge factorladingen vaak vet gedrukt en de lage factorladingen dun gedrukt.
- Bij scheve rotatie moet je ook kijken naar de sterkte en het teken van de correlaties tussen de dimensies.
- Je kunt in SPSS niet de richting van de ladingen (negatief of positief teken) beheersen: soms kom je een hele serie negatieve ladingen tegen, die je mentaal even moet ompolen, evenals de betrokken correlaties tussen de betrokken dimensies.

Verschil tussen factor- en componentenanalyse

- Factoranalyse veronderstelt dat er sprake is van een of meerdere *latente* variabelen die de (cor)relaties tussen de indicatoren verklaren. Dit is de meest voorkomende interpretatie, ook bij toepassingen van componentenanalyse. Bij componentenanalyse gaat het er feitelijk om welke index het meest onderscheid tussen cases aanbrengt. Componenten zijn geen latente variabelen, je kunt ze laten berekenen, ze worden dan een kolom in je data-matrix.

- Componentenanalyse is rekenkundig stabiel; bij latente factoranalyse kan het rekentuig wel eens op hol slaan (zeldzaam).
- Met name bij latente factoranalyse ben je ook inhoudelijk erg geïnteresseerd in de correlaties tussen de latente factoren: deze zijn een schatting van de correlaties tussen latente factoren, *gecorrigeerd voor meetonbetrouwbaarheid*. Deze zijn doorgaans hoger dan correlaties tussen (berekende) componenten.

Bijlage A: Minimale SPSS syntax (onderlijnd standaardwaarden)

```
FACTOR Var=varlist  
/Missing=pair / list  
/Print=uni corr extraction / rotation  
/Format=sort blank(0.30)  
/Extraction=PAF / PC  
/Criteria=factors(N)  
/Rotation=oblimin / varimax.
```

Bijlage B: Factoranalyse-termen en hun betekenis in meer gebruikelijke termen	
Factor	Latente, ongeobserveerde variabele; dimensie.
Loading	Direct effect van latente variabele op geobserveerde indicator; gestandaardiseerd
Extraction	Onbelangrijk, behalve dat je hiermee kiest tussen latente factoranalyse (PAF) en componentenanalyse (PC)
Communaliteit	Verklaarde variantie in afzonderlijke geobserveerde indicatoren.
Pattern matrix	Matrix met directe effecten van latente variabelen op geobserveerde indicatoren
Structure matrix	Matrix met totale (directe + indirecte) effecten van latente variabelen op geobserveerde indicatoren
Eigenvalue	Som van de verklaarde varianties in geobserveerde indicatoren per latente variabele.
Factor correlation	Correlatie tussen de latente variabelen
Oblique (scheve) rotatie	Het toestaan van correlatie tussen latente factoren (een heel natuurlijke veronderstelling)
Orthogonale (rechte) rotatie	Het niet toestaan van correlatie tussen latente factoren (een weinig natuurlijke veronderstelling die vaak tot minder simple structure leidt). In SPSS vraag je dit op met /rotation=varimax.
Simple structure, varimax	Eenvoudig patroon van factorladingen in de pattern matrix: zoveel mogelijk dichtbij 0 of 1.0. Simple structure zie je in de pattern matrix.
Component	Gewogen somscore van geobserveerde indicatoren. Bij multiële componentenanalyse zijn de achtereenvolgende componenten ongecorrleerd. Componentenanalyse is de standaard instelling van SPSS factoranalyse.