

STAPPENPLAN BETROUWBAARHEIDSANALYSE

Vooraf: Betrouwbaarheidsanalyse is een techniek om uit te maken hoe groot de betrouwbaarheid is van een index die is samengesteld uit ongewogen middeling van meerdere indicatoren.

- Betrouwbaarheid is een schatting van de stabiliteit van de index, dat wil zeggen de correlatie tussen een index en een nieuwe meting ervan indien het gemeten kenmerk niet veranderd is. Oftewel: de correlatie van de index met zichzelf. Betrouwbaarheid is een correlatie en varieert tussen 0 en 1.
- Deze correlatie is doorgaans niet 1.0, omdat de scores worden beïnvloed door random meetfouten (ruis), dat zijn meetfouten die tussen twee meetmomenten op basis van volstrekt toeval (loting) fluctueren.
- Een andere definitie van een betrouwbaarheidscoëfficiënt is de proportie ‘ware variantie’ in een (gestandaardiseerd) gemeten kenmerk.
- In de onderzoeksliteratuur wordt er vaak gezegd dat de betrouwbaarheid een bepaalde minimumgrens moet overschrijden, maar deze grens fluctueert nogal: 0.60, 0.65, 0.70, 0.75, 0.80 worden allemaal wel eens genoemd.
- Je kunt betrouwbaarheid ook zien als de mate van afzwakking (*attenuation*) die optreedt in correlatie- en regressiecoëfficiënten door random meetfouten. De mate van afzwakking is evenwel $1 - \sqrt{\text{betrouwbaarheid}}$: een betrouwbaarheid van 0.81 indiceert 10% afzwakking ($\sqrt{0.81} = 0.90$).
- Betrouwbaarheidsanalyse is niet bedoeld en niet geschikt om uit te maken of je indicatoren één dimensie vormen. Daarvoor gebruik je factor- of componentenanalyse.
- We hebben het hier over een index en indicatoren. Andere vaak gebruikte termen voor index zijn: schaal, construct, factor, component. Andere vaak gebruikte woorden voor indicatoren zijn: items, metingen.
- De betrouwbaarheid van een index wordt zowel vergroot door het aantal betrokken indicatoren als de (gemiddelde) correlatie tussen deze indicatoren. Je kunt betrouwbaarheid zowel vergroten door beter en als door vaker te meten. Betrouwbaarheidsanalyse in SPSS maakt een afweging tussen deze twee determinanten: bij meer indicatoren is de gemiddelde correlatie meestal wat lager.
- Betrouwbaarheidsanalyse veronderstelt dat je de index zonder weging van de indicatoren samenstelt. Je kunt een (optimale) gewogen constructie verkrijgen via PC (in factoranalyse). Het verschil meestal niet de moeite waard.
- **Stap 1:** Maak een frequentieverdeling van alle indicatoren en vergewis je ervan dat de poling begrijpelijk is.
 - Poling wordt bepaald door zowel de inhoud van de indicator als de richting van de antwoorden.
 - Voor interpretatie is het het gemakkelijkst als je een index zodanig kunt benoemen dat *positieve* antwoorden op *meer* aanwezigheid van het construct wijst (voor een correcte analyse is echter alleen noodzakelijk dat de indicatoren dezelfde kant op gepoold zijn).
- **Stap 2:** Bekijk via descriptives een compact overzicht van het bereik (min en max), gemiddelde en standaarddeviaties van de indicatoren.

- Als indicatoren op een verschillende meetschaal gemeten zijn (bv. een indicator is een 0/1 variabele en een andere gaat van 1 naar 5, dien je de indicatoren te standaardiseren tot Z-scores.
 - Ook als indicatoren op dezelfde meetschaal gemeten zijn, kan de feitelijke meetschaal uiteenlopen: de standaarddeviaties verschillen dan sterk. Ook in dat geval is standaardisatie tot Z-scores aan te bevelen.
 - Standaardisatie tot Z-scores gaat handig via **descriptives**.
 - De noodzaak tot standaardisatie wordt geïndiceerd door verschil in standardized en unstandardized alpha.
- **Stap 3:** Zijn er missing values? In dat geval zul je moeten beslissen wat je met de partieel gemeten cases wil doen.
 - Reliability (in SPSS) heeft alleen een *listwise* variant. Dit is enigszins paradoxaal, omdat de berekening van betrouwbaarheid alleen gebruik maakt van gemiddelde correlatie en het aantal indicatoren, hetgeen ook voor partieel gemeten cases gedefinieerd is.
 - Let er vooral op hoeveel cases er in totaal niet worden meegenomen door missing values (listwise N). Het kan voorkomen dat je analyse maar op een kleine subset van de gegevens betrekking heeft! In zo'n geval kan je via MVA de missende waarden substitueren.
 - Constructie van een index uit meerdere indicatoren is echter juist een probate manier om missende waarden te substitueren.
- **Stap 4:** Voer de betrouwbaarheidsanalyse uit.
 - Noteer de geschatte betrouwbaarheid van de index (alpha) en vergelijk deze met 'alpha if item deleted'. Je kunt de betrouwbaarheid van de index verbeteren door stapsgewijs indicatoren te verwijderen die de alpha verlagen.
 - Een informatieve kolom is verder de "corrected item total correlation": dit is de correlatie tussen een indicator en het gemiddelde van de andere indicatoren (betere naam zou zijn: item-rest correlatie). **Deze kolom geeft een directe indicatie van evt. inconsistente poling van de indicatoren.**
 - De overige kolommen in de SPSS output zijn niet zo informatief. De Mean kolom geeft aan wat het gemiddelde van de index zou zijn als je die via sommering (dus niet via middeling) zou construeren. De Variance kolom geeft de variantie van die gesommeerde index aan: merk op dat deze maximaal is bij de hoogst mogelijke betrouwbaarheid. De Multiple R is het kwadraat van de corrected item-total correlation.
- **Stap 5:** Construeer de schaal als het gemiddelde van de gekozen indicatoren.
 - Bereken het gemiddelde via `compute index=mean(indicatoren)`. Op die manier behoud je cases met partiële missing values.
 - Rapporteer behalve de betrouwbaarheid ook het gemiddelde, bereik (min, max) en standaarddeviatie van de geconstrueerde index.
 - Ook al is je index in dezelfde eenheid als de samenstellende indicatoren, de meeteenheid is toch niet identiek, omdat de variantie veranderd is. Merk op dat als je Z-gestandaardiseerde indicatoren middelt, het resultaat wel een gemiddelde 0 heeft, maar NIET een standaarddeviatie van 1. Als je dit nuttig vindt, moet je de schaal zelf opnieuw standaardiseren.

Bijlage A: Minimale SPSS syntax (onderlijnd voorkeurswaarden)

```
RELIability VAR=varlist  
/STATistics=Descriptives Correlations  
/SUMMarize=Total / All.
```

Bijlage B: Verdere toelichting

Betrouwbaarheidanalyse in SPSS berust op de *interne consistentie coefficient* α van Cronbach (1951). Uitgedrukt in inter-itemcorrelaties luidt deze:

$$\alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}}$$

k = aantal vragen/ items

\bar{r} = gemiddelde correlatie

Deze coëfficiënt doet een *schatting van* de betrouwbaarheid van een index op basis van de onderlinge samenhang van de betrokken indicatoren. De schatting berust onder meer op de volgende aannamen:

- De samenhang tussen de indicatoren wordt veroorzaakt door een enkele onderliggende latente variabele (factor).
- De betrokken indicatoren hebben (ongeveer) gelijke varianties en (ongeveer gelijke) factorladingen.

Voor dichotome (0/1) indicatoren wordt dezelfde coëfficiënt meestal aangeduid als KR20 (de 20^{ste} formules van Kuder & Richardson (1937)). Bij dichotome indicatoren zien we soms dat indicatoren met een kleine variantie (bijna iedereen in een van beide categorieën) gediskwalificeerd worden, hoewel ze in feite een goede meting zijn van de uiteinden van de schaal. In deze gevallen zijn anders schaalmodellen, zoals dat van Guttman, Mokken of Rasch meer geëigend.

Betrouwbaarheid is niet hetzelfde als de coëfficiënt van Cronbach, dat is slechts een schatting van betrouwbaarheid als je meerdere indicatoren hebt die volgens een factormodel samenhangen. Een meer algemeen geldige definitie van betrouwbaarheid is “de correlatie van een variabele met zichzelf” (test-retest betrouwbaarheid). Die definitie is ook van toepassing op een enkele indicator of een index met alternatieve metingen¹, maar kun je alleen gemakkelijk toepassen als het om onveranderlijke kenmerken gaat. Betrouwbaarheid van een meting van opleiding (en kenmerk dat niet verandert) is bv. hetzelfde als de correlatie tussen twee metingen van opleiding in opeenvolgende interviews. Bij veranderlijke kenmerken (bv. een mening) kun je betrouwbaarheid alleen via de test-retest methode berekenen als je mag veronderstellen dat het kenmerk niet veranderd is. Dat is soms het geval op zeer korte termijn (binnen hetzelfde interview).

¹ Denk bv. aan het meten van contactfrequentie tussen personen via de hoeveelheid persoonlijke gesprekken, telefoongesprekken en e-mails. Zulke alternatieve indicatoren zijn niet intern consistent en je kunt Cronbach's alpha daarom niet gebruiken.