

Onderzoek: betrouwbaarheid en validiteit



Bij elk wetenschappelijk verantwoord onderzoek moet de vraag worden beantwoord of het meetinstrument waarmee we gemeten hebben, betrouwbaar en valide is. Dat meetinstrument kan een vragenlijst zijn, maar bijvoorbeeld ook een liniaal of een weegschaal. Bij elke meting kunnen we fouten maken. Sommige fouten zijn toevallig, terwijl andere systematisch optreden.

Betrouwbaarheid en validiteit zijn checks op twee verschillende soorten fouten: betrouwbaarheid betreft een controle op toevallige fouten ('toevalsfouten') en validiteit verwijst naar een controle op systematische fouten. Toevalsfouten ontstaan vaak door de omgeving waarin je meet; de situatie is dan van invloed op het meetresultaat. Systematische fouten zijn een graad ernstiger; het meetinstrument meet dan in elk denkbare situatie niet wat het beoogt te meten.

Hieronder lichten we het verschil tussen toevals- en systematische fouten op een praktische wijze toe. Daarna beschrijven we per type fout de methoden die je kunt gebruiken om de vraag te beantwoorden of een meetinstrument betrouwbaar en/ of valide is.

Toevalsfouten

Stel dat we de lengte van een plastic trommel willen meten met een metalen liniaal. Als we dit met dezelfde liniaal eerst op de evenaar doen en later op de noordpool, dan zullen we zien dat de trommel op de evenaar kleiner is dan op de noordpool. Als we er van uitgaan dat de plastic trommel niet uitzet of krimpt onder verschillende weersomstandigheden, moet het dus logischer wijze aan de metalen liniaal liggen; op de evenaar zet de liniaal uit en op de noordpool krimpt deze in. Door in andere situaties te meten, krijgen we met deze liniaal dus verschillende resultaten. De liniaal geeft met andere woorden geen betrouwbaar resultaat over de lengte van de plastic trommel. Een geheel ander voorbeeld betreft de vraag welk bier je bij voorkeur koopt. Als je die vraag beantwoordt vanuit de gedachte dat je het bier koopt voor een avondje waarbij je vrienden op bezoek komen, zul je waarschijnlijk een heel ander merk noemen dan dat je het merk in gedachten neemt voor individuele consumptie. Ook hier beïnvloedt de situatie de uitkomst van het onderzoek en spreken we over *toevalsfouten*.

Systematische fouten

Bij systematische fouten is er wat anders aan de hand dan bij toevalsfouten. Laten we weer even teruggaan naar de liniaal waar we het hierboven over hadden. Stel dat dit een *plastic* liniaal is die *geen* last heeft van temperatuursinvloeden. Maar, bij deze liniaal is de schaal van 40 centimeter die er op afgedrukt staat, door een productiefout opgerekt. De schaal van 40 centimeter is in werkelijkheid 50 centimeter. Waar we nu ook meten, deze liniaal resulteert altijd – systematisch – in een foute uitkomst. In vragenlijsten kunnen systematische fouten voorkomen als we de verkeerde vraag stellen en daardoor niet meten wat we eigenlijk willen meten. Een voorbeeld daarvan is als we de vraag willen beantwoorden of mensen regelmatig bier *drinken* en hen de vraag stellen of ze regelmatig bier *kopen*. Uiteraard kan het dan zo zijn dat een respondent aangeeft dat hij/ zij regelmatig bier koopt, maar het zelf niet drinkt (en het dus voor bijvoorbeeld een huisgenoot koopt). Ongeacht de situatie zul je met deze vraag nooit meten of mensen regelmatig bier drinken.

Relevante vraag is hoe je kunt nagaan of een onderzoek geen toevals- of systematische fouten bevat. Wetenschappers hebben daarvoor een aantal checks bedacht. Hieronder bespreken we eerst controle-instrumenten voor toevalsfouten (betrouwbaarheid) en daarna die voor systematische fouten (validiteit).

Checks op betrouwbaarheid

Er zijn vier checks voor betrouwbaarheid, ofwel instrumenten voor de controle op toevalsfouten:

1. Item-analyse (homogeniteit): deze methode kijkt hoe homogeen een aantal vragen in een vragenlijst hetzelfde construct meet. Aangezien in een vragenlijst vaak meerdere constructen worden gemeten, is het noodzakelijk van te voren te bepalen op welke vragen deze analyse wordt toegepast. Door in de analyse bepaalde items (vragen) weg te laten, wordt nagegaan hoe homogeen de score is. De analysemethode die hiervoor wordt gebruikt is Cronbach α .
2. Split half (interne consistentie): een (deel van een) vragenlijst waarin de vragen over eenzelfde construct gaan, wordt opgesplitst in tweeën. De correlatie tussen beide delen geeft aan hoe betrouwbaar de meting is.
3. Test - retest (stabiliteit): een onderzoek wordt twee keer afgenomen; de correlatie tussen beide onderzoeken geeft aan hoe betrouwbaar het onderzoek is.
4. Parallel test (a-specificiteit): naast het eigen meetinstrument wordt nog een zelfde meting gedaan met een ander meetinstrument dat wel hetzelfde beoogt te meten.

Checks op validiteit

Naast het bepalen van de betrouwbaarheid van een onderzoeksinstrument, moet ook de validiteit ervan worden bepaald. Een instrument dat niet betrouwbaar is, kan *niet* valide zijn; 'betrouwbaarheid is een noodzakelijke maar niet voldoende voorwaarde voor validiteit'. Er zijn drie hoofdvormen van validiteit te onderscheiden, met bijbehorende checks:

1. Inhouds/ content validiteit: is de meting van de variabele een goede afspiegeling van alle denkbare situaties? Deze check kan uitgevoerd worden door verschillende situaties te benoemen en te beredeneren waarom de meting in al die situaties meet wat het wordt geacht te meten.
2. Begrips/ construct validiteit: wijze van logisch redeneren waarbij de relatie tussen het eigen meetinstrument en andere meetinstrumenten centraal staat:
 - a. convergente validiteit: deze vorm van validiteit kan worden bepaald door de eigen meetmethode qua opzet te vergelijken met andere meetmethoden die geacht worden *hetzelfde* te meten;
 - b. discriminant validiteit: onderscheidendheid van het eigen meetinstrument ten aanzien van andere meetmethoden die geacht worden *andere* constructen te meten;
 - c. nomologische validiteit: het op een logische wijze kunnen plaatsen van variabelen uit het eigen meetinstrument in een theoretisch model (een 'nomologisch netwerk').
3. Criterium validiteit (relatie van een variabele met een extern criterium):
 - a. concurrente validiteit: correspondentie met gelijktijdig beschikbare criteriumgegevens (de uitslag van een Cito-toets wordt vergeleken met gelijktijdig verkregen rapportcijfers);
 - b. predictieve validiteit: correspondentie met in de toekomst gelegen criteriumgegevens (de uitslag van een Cito-toets wordt vergeleken met resultaten die later op de middelbare school worden gehaald).