

# 1. Waarom dwingt mijn duivelse docent me tot statistiek?

## Methoden

Onderzoeken waarbij gebruik gemaakt van data op basis van cijfers worden *kwantitatieve methoden* genoemd en onderzoeken waarbij taal wordt gebruikt als basis voor het onderzoek, worden *kwalitatieve methoden* genoemd. Beide methoden zijn complementair aan elkaar, de een is niet beter dan de ander.

## De observatie

Het onderzoeksproces bestaat uit een aantal stappen. De eerste stap is observatie. Hierbij neem je iets waar wat je nieuwsgierig maakt, er komt een vraag in je op die je graag beantwoord wilt hebben. Om te kijken of de observatie klopt moet je gegevens verzamelen. Voor het verzamelen van deze gegevens heb je *variabelen* nodig. De variabele is hetgeen dat je meet om tot een antwoord op je vraag te komen.

## Verklaren van de gegevens

De volgende stap in het onderzoeksproces is het verklaren van de data. Dit doe je met behulp van een *theorie*. Op basis van een theorie kan je een voorspelling doen. Deze voorspelling op basis van een theorie noem je een *hypothese*. Om te spreken van een hypothese moet het gaan om een uitspraak die je met behulp van wetenschappelijke methodes kunt bewijzen of verwerpen. Wanneer de verzamelde gegevens de theorie of hypothese tegenspreken, is er sprake van *falsificatie*.

## Gegevens verzamelen

Als we gegevens willen verzamelen is het belangrijk dat we ons twee dingen afvragen: (1) wat wordt er gemeten en (2) hoe wordt het gemeten? Om de hypothesen te kunnen testen moeten we variabelen meten. Variabelen zijn dingen die kunnen variëren, tussen mensen, of tussen situaties of in de loop van de tijd. Bij de meeste hypothesen heb je twee variabelen, de oorzaak en de uitkomst. De variabele die we zien als de oorzaak wordt de *onafhankelijke variabele* of *predictor* genoemd. Bij een experimentele opzet wordt deze term gebruikt om te benadrukken dat de onderzoeker deze variabele gemanipuleerd heeft. De variabele die verandert door veranderingen in de onafhankelijke variabele wordt de *afhankelijke variabele* of *uitkomst variabele* genoemd.

## Meetniveaus

Variabelen kunnen op verschillende manieren gemeten worden. Het verband tussen wat je meet en de cijfers die uitdrukken wat je meet, noem je het *meetniveau*. Variabelen kunnen categorisch of continu zijn.

Een *categorische variabele* bestaat uit verschillende categorieën. Je kan maar in één categorie per keer ingedeeld worden, je hoort niet een beetje bij de ene categorie en een beetje bij de andere categorie. Een voorbeeld van een categorische variabele is de indeling tussen mannen en vrouwen. In dit geval heeft de variabele slechts twee categorieën, waarbij je slecht in één van beide kunt passen. Een variabele met twee categorieën noem je een *binair variabele*.

Als een variabele bestaat uit meer dan twee categorieën die aan elkaar gelijk zijn, noem je dit een *nominale variabele*. Een voorbeeld van een nominale variabele is religie (Jodendom, Christendom, Islam, et cetera). Hoewel deze categorieën ook kunnen worden weergegeven met cijfers, is het niet mogelijk om wiskundige berekeningen uit te voeren met deze cijfers. Deze cijfers geven namelijk niet een rangorde aan bij een nominale variabele. Een voorbeeld van een nominale variabele die wordt weergegeven met cijfers is het rugnummer van een speler in een teamsport. Een hoger rugnummer betekent niet dat iemand een betere speler is. Nominale data kan alleen gebruikt worden om te kijken naar frequenties, dus hoe vaak een bepaalde speler scoort, of hoeveel mensen een bepaald geloof hebben.

Bij een *ordinale variabele* heb je ook verschillende categorieën, maar hebben die een bepaalde rangorde. Ordinale data geven bijvoorbeeld een bepaalde volgorde aan. Echter wordt er niet aangegeven hoe groot het verschil is tussen de categorieën. Een top drie van een wedstrijd geeft aan wie beter is dan de ander, dus het heeft een volgorde, maar het zegt niet hoeveel beter de winnaar was dan de nummer twee en drie.

Bij het volgende meetniveau heb je geen categorische variabele meer, maar is de variabele continu. Een *continue variabele* is een score die elke waarde kan aannemen die op de meetschaal wordt gebruikt. De *intervalvariabele* is een vorm van een continue variabele. Bij de intervalvariabele is het verschil tussen alle getallen gelijk. Een voorbeeld hiervan is een schaal waarbij je aangeeft hoe aardig je iemand vindt op een vijfpuntsschaal. Het verschil tussen 1 en 2 is hierbij even groot als het verschil tussen 4 en 5. Dit meetniveau wordt het meest gebruikt bij statistische testen.

Nog een stap verder is de *ratio variabele*. De ratio variabele heeft dezelfde voorwaarden als de intervalvariabele, alleen heeft de ratio variabele een absoluut en betekenisvol nulpunt. Dit betekent dat je de getallen van een ratio variabele kan vermenigvuldigen. Een voorbeeld hiervan is reactietijd. Een milliseconde duurt altijd even lang, dus de verschillen tussen de milliseconden zijn gelijk, maar daarnaast kun je ook zeggen dat 200 milliseconden twee keer zo lang is als 100 milliseconden.

Een continue variabele hoeft niet altijd continu te zijn, het kan ook een *discrete variabele* zijn. Een echte continue variabele kan alle mogelijke waarden aannemen, maar bij een discrete variabele kunnen alleen bepaalde waarden (meestal alleen hele getallen) gekozen worden. Als je aangeeft hoe aardig je iemand vindt op een vijfpuntsschaal, is dat wel een continuüm, waarbij 2.98 een betekenisvolle waarde is, maar je kunt alleen de cijfers 1, 2, 3, 4 en 5 werkelijk kiezen. Je kunt niet daadwerkelijk 2.98 invullen.

## Meetfouten

We willen het liefst een meting die hetzelfde is over de tijd en over situaties. Je wil een accurate meting, die niet beïnvloed wordt door wie of waar de meting wordt gedaan. Er is vaak een verschil tussen de gemeten waarde en de werkelijke waarde. Dit verschil noem je de *meetfout*. Als je een goed instrument hebt, is de meetfout klein. Bij vragenlijsten over gevoelige onderwerpen heb je vaak een grotere meetfout, omdat niet alleen de werkelijke situatie de antwoorden van de participanten beïnvloedt, maar er ook andere factoren zoals sociale wenselijkheid meespelen.

## Validiteit en betrouwbaarheid

Een manier om de meetfout zo klein mogelijk te maken is door eigenschappen van het meetinstrument vast te stellen die iets zeggen over hoe goed het meetinstrument zijn werk doet. Een manier om dat te bepalen is *validiteit*.

De validiteit betekent of het instrument daadwerkelijk meet wat je wilde meten. *Criterion validiteit* betekent dat je kunt vaststellen of je instrument meet wat je wil meten aan de hand van objectieve criteria.

Dit kan op twee manieren gebeuren. Als je tegelijkertijd data verzamelt met het nieuwe instrument en bestaande criteria toetst, meet je de gelijktijdige validiteit. Als je de data van je nieuwe instrument gebruikt om latere observaties te voorspellen, meet je voorspellende validiteit. Het probleem met criterium validiteit is dat het niet altijd gebruikt kan worden omdat er vaak geen objectieve criteria bestaan voor wat je wil meten, zoals wanneer je wil weten hoe aardig iemand wordt gevonden.

Een andere vorm van validiteit is *inhoudsvaliditeit*. Dit gaat over de mate waarin de items op een vragenlijst overeenkomen met het construct en of de vragen het fenomeen volledig dekken.

Een instrument moet valide zijn, maar dat is nog niet voldoende. Het moet daarnaast ook *betrouwbaar* zijn. Betrouwbaarheid betekent dat het instrument dezelfde uitkomst geeft bij dezelfde condities. Dus een betrouwbare weegschaal geeft altijd hetzelfde gewicht als het werkelijke gewicht hetzelfde is. Als een instrument niet betrouwbaar is, kan het ook niet valide zijn. Een instrument dat onder dezelfde omstandigheden namelijk verschillende uitkomsten genereert, meet per definitie niet wat het hoort te meten. De makkelijkste manier om de betrouwbaarheid te testen is door de test te herhalen (*test-hersteet betrouwbaarheid*). Een betrouwbaar instrument geeft dan op beide testen een vergelijkbaar resultaat.

## Verschillende onderzoeksmethoden

Er zijn grofweg twee manieren om data te verzamelen, met *correlatieve onderzoek* (*cross-sectioneel onderzoek*) en *experimenteel onderzoek*. Bij correlatieve onderzoek observeren we wat er in de wereld gebeurt zonder het te manipuleren. Dit is goed voor de *ecologische validiteit* omdat men de natuurlijke situatie observeert. Sommige onderzoeken kunnen alleen op deze correlatieve manier worden uitgevoerd, omdat het onmogelijk of onethisch is om bepaalde variabelen te manipuleren. Het nadeel van deze methode is echter dat het niet mogelijk is om een uitspraak te doen over causaliteit.

Bij experimenteel onderzoek wordt een variabele gemanipuleerd om te kijken of het de andere variabelen beïnvloedt. Bij veel onderzoeken wordt gekeken of de ene variabele (de predictor) de oorzaak is van de andere variabele (de afhankelijke variabele). Men kan volgens Hume pas spreken van een oorzakelijk verband als:

- oorzaak en gevolg elkaar nauw in de tijd opvolgen
- de oorzaak voorafgaat aan het gevolg
- het gevolg nooit voorkomt zonder dat de oorzaak heeft plaatsgevonden.

Bij veel onderzoeken worden de variabelen tegelijkertijd gemeten. In veel gevallen weet men dan niet welke variabele de oorzaak is en welke variabele het gevolg weergeeft. Het kan zijn dat er een derde variabele in het spel is (*tertium quid*) die de oorzaak is van beide andere variabelen. Dit wordt ook wel de *verwarrende variabele* genoemd. Een voorbeeld is het verband tussen borstimplantaten en zelfmoord. Een laag zelfbeeld is de oorzaak van het nemen van een borstvergroting en het ondernemen van zelfmoordpogingen.

John Stuart Mill heeft nog een criterium toegevoegd aan de criteria van Hume, namelijk dat alle andere verklaringen van het oorzaak-gevolg effect uitgesloten moeten zijn. Wanneer de oorzaak afwezig is, mag het gevolg ook niet aanwezig zijn. Bij experimenteel onderzoek is het doel zo goed mogelijk de oorzaak-gevolg relatie tussen variabelen te vinden. Experimenten vergelijken situaties (*condities*) waarin de vermeende oorzaak afwezig is met de conditie waarbij de oorzaak aanwezig is.

De deelnemers kunnen op twee verschillende manieren meedoen aan een experiment. Dit kan met een *binnen-groep ontwerp* en in een *tussen-groep ontwerp*.

Bij een binnen-groep ontwerp doen dezelfde deelnemers het experiment een aantal keer in verschillende condities. Bij een tussen-groep ontwerp heb je verschillende deelnemers in de verschillende condities.

*Niet-systematische variatie* is het kleine verschil in prestatie tussen twee condities dat niet door de bekende factoren verklaard kan worden. Zelfs als alle variabelen gelijk blijven, is er alsnog meestal een klein verschil in scores tussen verschillende condities of verschillende momenten. Dit komt bijvoorbeeld door verschillen in vaardigheid in de taak tussen mensen of verschillende tijdstippen van de dag.

*Systematische variatie* is het verschil tussen de twee condities die ontstaat door de manipulatie van de conditie. Bijvoorbeeld, in de ene conditie krijgen chimpansees een beloning voor hun gedrag, en in de andere conditie niet. Het verschil in gedrag wordt nu veroorzaakt door de manipulatie, dit is de systematische variatie.

De rol van de statistiek is om te ontdekken hoeveel verschil er is in prestatie en welk deel van de variatie dan systematisch is en welk deel niet systematisch is. Bij het binnen-groep ontwerp is er minder niet-systematische variatie dan bij het tussen-groep ontwerp. Bij het tussen-groep ontwerp kunnen de personen in de verschillende groepen namelijk andere eigenschappen hebben. Om de niet-systematische variatie zo klein mogelijk te houden en de test zo nauwkeurig mogelijk te maken, maken wetenschappers gebruik van randomisatie.

Randomisatie is belangrijk omdat andere bronnen van systematische variatie verwijderd worden, waardoor we zeker weten dat veranderingen komen door de experimentele manipulatie. In het binnen-groep ontwerp zijn er nog twee belangrijke bronnen van systematische variatie:

*Oefeneffecten*: deelnemers kunnen zich de tweede keer anders gedragen bij de test, omdat ze bekend zijn geworden met de test.

*Vervelingseffecten*: deelnemers kunnen zich in de tweede test anders gedragen, omdat ze door de eerste test verveeld of moe zijn geworden.

Om deze effecten zo klein mogelijk te houden worden er bij verschillende deelnemers verschillende volgordes van conditie gegeven (*counterbalancing*), dus de ene persoon krijgt eerst conditie 1 en dan 2, en de ander eerst conditie 2 en dan 1. Wie welke conditie het eerst krijgt wordt gerandomiseerd, dus willekeurig toegewezen.

In het tussen-groep ontwerp wordt er gerandomiseerd door de deelnemers willekeurig toe te wijzen aan de verschillende condities. Mensen verschillen immers in eigenschappen, die mogelijke confounders (verwarrende variabelen) zijn. Als de participanten at random verdeeld zijn over de condities, is deze variatie deel van de niet-systematische variatie. De groepen verschillen dan niet op een systematische manier van elkaar, anders dan in de experimentele manipulatie.

## Het analyseren van de gegevens

Zodra je alle gegevens verzameld hebt, ga je de data analyseren. Hierbij is het handig om er een grafische weergave van te maken. Dit kan met een *frequentieverdeling* (ook wel een *histogram* genoemd). In deze grafiek is te zien hoe vaak een bepaalde score voorkomt in je data. De grafiek is handig bij het uitrekenen van de proporties.

Als je een verticale lijn door het midden van een histogram trekt en de beide helften zijn symmetrisch, dan heb je de ideale situatie. Dit wordt een *normale verdeling* genoemd. Een normale verdeling is een belvormige curve. Dit betekent dat de meeste scores rond het midden van de verdeling zitten. Veel fenomenen zijn normaal verdeeld.

Wanneer een histogram niet symmetrisch is loopt het *scheef*. Als het histogram veel scores aan de linkerzijde heeft is het positief scheef en als het veel scores aan de rechterzijde heeft is het negatief scheef. *Kurtosis* geeft de mate aan waarin de scores zich in de staarten van de verdeling bevinden. Dit is te zien aan hoe puntig het histogram is. Bij een *leptokurtic verdeling* is de kurtosis positief en loopt het histogram in een puntige grafiek. Bij een *platykurtic verdeling* is de kurtosis negatief en is het histogram vlakker dan normaal.

Men kan berekenen waar het centrum van de frequentie verdeling ligt (*central tendency*). De simpelste methode hiervoor is de *modus*, dit is simpelweg de score met de hoogste frequentie, dus die het vaakst voorkomt. Er kunnen meerdere scores zijn met een gelijke frequentie. Als er twee scores zijn die het meest voorkomen, is de verdeling *bimodaal*. Bij meer dan twee modes is de verdeling *multimodaal*.

Een andere central tendency is de *mediaan*. De mediaan is middelste score wanneer je alle scores qua frequentie op volgorde van klein naar groot neer zet. De positie van de mediaan kan berekend worden met de formule:  $(n+1)/2$ . Als je 11 scores hebt, is de mediaan dus het zesde cijfer. Als je een even aantal scores hebt, valt je mediaan tussen twee scores. Dan bereken je de mediaan door het gemiddelde van die twee scores te nemen. Extreme scores en een scheve verdeling hebben weinig invloed op de mediaan. De mediaan kun je gebruiken met data op ordinaal, interval en ratio meetniveau. Voor nominale data is het niet te gebruiken, omdat je die niet op volgorde van klein naar groot kunt zetten.

De derde manier om het midden van de verdeling te berekenen is met het *gemiddelde*. Het gemiddelde bereken je door alle scores op te tellen en te delen door het aantal participanten. In formulevorm:

$$\bar{X} = \frac{\sum x}{n}$$

= gemiddelde

$\Sigma$  = somteken, het optellen van alle scores

x = de score van een deelnemer

n = aantal deelnemers, steekproefgrootte

Het gemiddelde kan beïnvloed worden door extreme scores en door een scheve verdeling. Ook is het alleen te gebruiken bij interval of ratio gegevens. Het voordeel van het gemiddelde boven de mediaan en modus is dat je bij het berekenen ervan alle scores meeneemt. Bij de mediaan en modus neger je de meeste scores in je dataset.

Naast het midden van de verdeling, kun je ook geïnteresseerd zijn in hoe de scores verspreid zijn. De *range* (reikwijdte) van de scores is de hoogste score min de laagste score. Omdat je alleen de hoogste en laagste score gebruikt om de range te berekenen, hebben extreme scores veel invloed. Om deze invloed te verminderen wordt de bovenste 25% en de onderste 25% weggehaald waardoor je de middelste 50% overhoudt. Dit wordt de *interkwartielafstand* genoemd. Kwartielen zijn de drie waarden die de verdeling in vier gelijke stukken verdelen. De mediaan van de data is het *tweede kwartiel*. Het eerste kwartiel is de mediaan van de laagste helft, het derde kwartiel is de mediaan van de bovenste helft.

Het nadeel van de range is dat je de helft van de data niet gebruikt. Als je wel alle gegevens wil gebruiken, kun je kijken hoe ver elke score verwijderd ligt van het midden van de verdeling. Deze *afwijking* of *deviatie* bereken je met: score – gemiddelde. De totale deviatie bereken je door alle deviaties op te tellen.

Omdat sommige scores boven het gemiddelde liggen, en andere eronder, kom je bij de totale deviatie altijd uit op 0. Daarom tellen we de deviatiescores niet gewoon op, maar worden de waarden eerst gekwadrateerd en dan opgeteld. Deze opgetelde kwadraten noemen we de *kwadratensom* of *sum of squares (SS)*. Negatieve waarden worden positief in het kwadraat, dus de SS is altijd groter dan nul (of nul, wanneer alle scores exact gelijk zijn).

Het probleem met deze kwadratensom is dat de grootte ervan afhangt van het aantal scores dat je optelt. Hierdoor kun je kwadratensommen van verschillende steekproefgroottes niet vergelijken. Daarom wordt er een gemiddelde spreidingsmaat gebruikt, die we de *variantie* noemen. De variantie ( $s^2$ ) is de kwadratensom gedeeld door de steekproefgrootte  $- 1$ .

Het nadeel van de variantie is dat het een gekwadrateerde maat is. Meestal wordt daarom de wortel van de variantie gebruikt, de *standaardafwijking* ( $s$ ).

De kwadratensom, de variantie en de standaardafwijking zijn allemaal maten die aangeven hoe ver de data verspreid is rond het gemiddelde. Een grote standaardafwijking betekent een grote spreiding, waarbij er veel scores ver van het gemiddelde zitten, en bij een kleine standaardafwijking zitten alle scores ongeveer rond het gemiddelde. Bij een grote standaardafwijking wordt de verdeling platter, bij een kleine standaardafwijking is de verdeling puntiger. Dit kan lijken op een platykurtic of leptokurtic verdeling, terwijl dat niet zo is.

Je kunt frequentieverdelingen niet alleen gebruiken om te bekijken hoe vaak bepaalde scores daadwerkelijk voorkwamen, maar ook om een uitspraak te doen over hoe waarschijnlijk het is dat iets voorkomt. Als er 172 zelfmoorden waren, waarvan 36 personen tussen de 30 en 35, betekent dit een proportie van  $36/172=0.21$ , ofwel 21%. Met deze proporties kun je inschatten hoe groot de kans is dat een bepaalde score voorkomt. Kansen nemen een waarde aan tussen de 0 (er is geen kans dat het gebeurt) en de 1 (het gaat zeker gebeuren).

Om deze kansen te berekenen, maak je gebruik van een *kansverdeling*. Het gebied onder een deel van de kansverdeling geeft aan hoe groot de kans is dat een bepaalde waarde wordt verkregen. Vaak wordt er bij de normaalverdeling gebruik gemaakt van een standaardverdeling (*z-verdeling*), waarbij het gemiddelde altijd 0 is en de standaardafwijking altijd 1. Alle data sets kunnen worden omgezet in zo'n standaardverdeling. Dit doe je door de scores te veranderen in *z-scores*. Dit kan door de score  $-$  het gemiddelde te doen, en vervolgens te delen door de standaardafwijking.

Deze z-scores kunnen gebruikt worden om in de tabel van de standaard normaalverdeling de bijbehorende proporties op te zoeken. De proportie in de standaard normaalcurve is hetzelfde als de kans op die waarde. Een z-score van 2.6 betekent een score die 2.6 standaardafwijkingen boven het gemiddelde ligt. In de tabel (achter in het boek) is te zien dat een z-score van 2.6 correspondeert met een proportie in de normaalcurve van 0.0044. Dat betekent dat er 0.44% kans is op deze waarde. Het betekent ook dat 99.56% van de curve onder deze waarde ligt, want de hele curve is 1 (100%).

De tabel met de standaard normaalverdeling kun je ook gebruiken om de vraag te beantwoorden welke range de middelste 95% heeft. Omdat de normaalcurve symmetrisch is, betekent dit dat er 2.5% van de distributie aan beide kanten van het midden zit. Om erachter te komen welke z-score hoort bij de proportie 0.025, zoek je deze waarde in de tabel bij de kolom 'smaller portion'. De bijbehorende z-score is 1.96. Omdat de verdeling symmetrisch is, is  $-1.96$  de andere grens. De middelste 95% van de curve zit dus tussen de z-scores  $-1.96$  en  $1.96$ .

Nadat je onderzoek is afgerond, zul je een rapport moeten schrijven met je bevindingen en dat insturen naar een tijdschrift. Voor het rapporteren worden meestal de regels en richtlijnen van de American Psychological Association (APA) gevolgd.