

Statistiek en Meten: Wat moet je daarover weten?

Onder redactie van Diana Dolmans, Cees van der Vleuten, Albert Scherpbier en Ineke Wolfhagen

Bij het lezen van publikaties over onderzoek van onderwijs worden lezers regelmatig geconfronteerd met allerlei statistische begrippen, waarvan de betekenis niet geheel duidelijk is. Daarom heeft de redactie besloten om een reeks artikelen te publiceren waarin belangrijke onderwerpen op het gebied van statistiek en meten aan de orde worden gesteld. De redactie van deze reeks wordt gevormd door twee gastredacteuren (Diana Dolmans en Cees van der Vleuten) en twee leden van de BMO-redactie (Albert Scherpbier en Ineke Wolfhagen).

De keuze van onderwerpen is gebaseerd op veel gebruikte statistische technieken en begrippen. De nadruk ligt op de betekenis en interpretatie hiervan. Op deze manier wordt getracht een bijdrage te leveren aan de verdere professionalisering van docenten binnen het medisch onderwijs. Deze reeks start met de basale beginselen van de beschrijvende statistiek.

Beschrijvende Statistiek

D. Dolmans

Termen die aan bod komen:

Variabele, steekproef, populatie, meetniveau, absolute frequentieverdeling, relatieve frequentieverdeling, cumulatieve frequentieverdeling, normaalverdeling, rechtsscheve verdeling, linksscheve verdeling, maten van centrale tendentie (gemiddelde, modus en mediaan), range, standaarddeviatie en standard error

Beschrijvende statistiek houdt zich bezig met het meten, ordenen en systematisch weergeven van variabelen. Stel dat men wil onderzoeken op welke leeftijd vrouwen in Nederland hun eerste kind krijgen. Hiervoor moeten eerst gegevens verzameld worden. Datgene wat we in een onderzoek willen meten of manipuleren, heet *variabele*. In dit voorbeeld is leeftijd de variabele. In een dergelijk onderzoek wordt meestal niet de gehele *populatie* - in dit geval alle vrouwen in Nederland met (een) kind(eren) - betrokken, maar wordt een be-

perkte groep vrouwen geselecteerd van de totale populatie. De geselecteerde groep vrouwen noemen we de *steekproef* (N). Om inzicht te krijgen in de verzamelde gegevens maken we gebruik van beschrijvende statistiek. Hierbij worden gegevens geordend aan de hand van frequentietabellen en weergegeven in de vorm van maten van centrale tendentie (het gemiddelde, de mediaan en de modus) en spreiding (range en standaarddeviatie). In dit artikel ligt de nadruk op het interpreteren van de genoemde maten.

Meetniveaus

De manier waarop gegevens geordend en gepresenteerd kunnen worden is afhankelijk van de verschillende waarden die een variabele kan aannemen. In de statistiek spreken we van *meetniveau*. Alvorens gegevens geordend en weergegeven worden, dient het meetniveau

Meetniveaus	Kenmerken van meetniveaus			
	Kwalitatief onderscheid	Rangordening	Afstand gedefiniëerd	Absoluut nulpunt
Nominaal	x			
Ordinaal	x	x		
Interval	x	x	x	
Ratio	x	x	x	x

Figuur 1. Een overzicht van de verschillende meetniveaus en hun kenmerken

bepaald te worden. Er worden vier meetniveaus onderscheiden: nominaal meetniveau, ordinaal meetniveau, interval meetniveau en ratio meetniveau. In figuur 1 worden de kenmerken van deze meetniveaus schematisch weergegeven.

De waarden die een variabele van nominaal meetniveau aanneemt, zijn alleen kwalitatief te onderscheiden, sluiten elkaar uit en er is geen rangordening mogelijk. Voorbeelden van variabelen van nominaal niveau zijn geslacht en nationaliteit. Bij variabelen van ordinaalniveau is wel een rangordening mogelijk. Voorbeelden hiervan zijn hoge, midden en lage inkomsten en opleidingsniveau. Variabelen van intervalniveau onderscheiden zich van de twee eerder genoemde meetniveaus doordat ze niet alleen geordend kunnen worden, maar doordat de afstanden tussen twee opeenvolgende waarden gelijk zijn. Bovendien is er sprake van een willekeurig gekozen nulpunt. Voorbeelden van variabelen van intervalniveau zijn tijdsaanduiding en temperatuur. Een temperatuur van 20 graden celsius is twee maal zo warm als een temperatuur van 10 graden celsius, met andere woorden de afstanden tussen twee opeenvolgende waarden zijn gelijk. Daarnaast is een temperatuur van nul graden celsius een willekeurig gekozen nulpunt, omdat nul graden celsius niet betekent dat er geen temperatuur is. Ten slotte, variabelen van rationiveau onderscheiden zich van de eerder

genoemde doordat er sprake is van een absoluut nulpunt. Voorbeelden van variabelen van rationiveau zijn gewicht en leeftijd.

Het bepalen van het meetniveau is van groot belang, omdat het gebruik van bepaalde statistische maten afhankelijk is van het meetniveau. Bijvoorbeeld, het berekenen van een gemiddelde, een van de maten van centrale tendentie, is niet zinvol voor een variabele van nominaal niveau. Een gemiddelde score voor de variabele geslacht van 1.3 is immers niet interpreteerbaar.

Alvorens maten van centrale tendentie te berekenen is het van belang de gegevens eerst te ordenen aan de hand van frequentieverdelingen.

Frequentieverdelingen

Het begrip frequentieverdeling zal aan de hand van een voorbeeld geïllustreerd worden. Stel dat de eindscore op een toets bestaat uit het aantal goed beantwoorde vragen. Met behulp van een frequentieverdeling kunnen de verschillende eindscores die op de toets behaald zijn, weergegeven worden. Uit kolom 1 van tabel 1 blijkt dat het aantal juist beantwoorde vragen 3, 4, 6, 7, 8 en 9 bedraagt. Een *absolute frequentieverdeling* geeft aan hoe vaak elk van deze eindscores voorkomt. Uit kolom 2 van tabel 1 blijkt dat één persoon drie vragen goed beantwoord heeft, één persoon vier vragen, twee personen zes vragen, etcetera.

Tabel 1. Een voorbeeld van een absolute, relatieve en cumulatieve frequentieverdeling

Aantal goed beantwoorde toetsvragen	Absolute frequentie	Relatieve frequentie	Cumulatieve frequentie
3	1	10%	10%
4	1	10%	20%
6	2	20%	40%
7	3	30%	70%
8	2	20%	90%
9	1	10%	100%

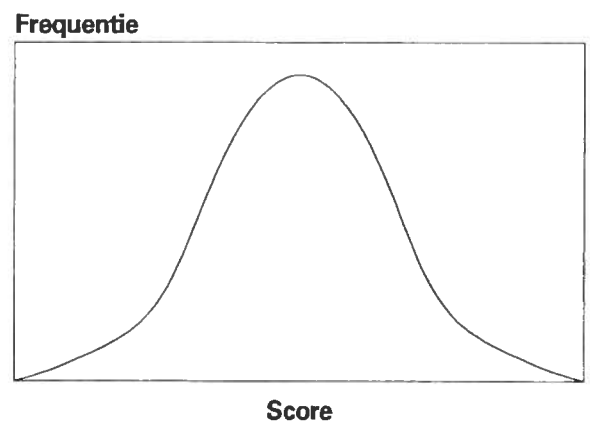
Een frequentieverdeling kan ook uitgedrukt worden in de vorm van percentages, dit noemen we een *relatieve frequentieverdeling* (tabel 1, kolom 3). Indien we de verdeling van de eindscores van mannelijke en vrouwelijke studenten op een toets willen vergelijken, maar twee groepen van verschillende grootte hebben, is het met behulp van een relatieve frequentieverdeling toch mogelijk beide verdelingen te vergelijken. Daarnaast worden soms *cumulatieve frequentieverdelingen* weergegeven. Een dergelijke verdeling geeft aan hoeveel procent van de totale steekproef beneden een bepaalde waarde scoort. Uit tabel 1, kolom 4, blijkt dat 70 procent van de personen een score heeft die gelijk is aan zeven goed beantwoorde vragen of minder.

Het voordeel van een cumulatieve frequentieverdeling zal geïllustreerd worden aan de hand van een voorbeeld. Indien een patiënt aan de arts vraagt of zijn of haar lengte normaal is, kan de arts aan de patiënt meedelen dat de gemiddelde lengte op de betreffende leeftijd 1.68 meter bedraagt. De arts kan echter ook meedelen dat 20 procent van de mensen uit die leeftijdscategorie kleiner is dan de patiënt zelf. De cumulatieve frequentieverdeling is echter alleen zinvol bij variabelen van ordinaal-, interval- of rationiveau.

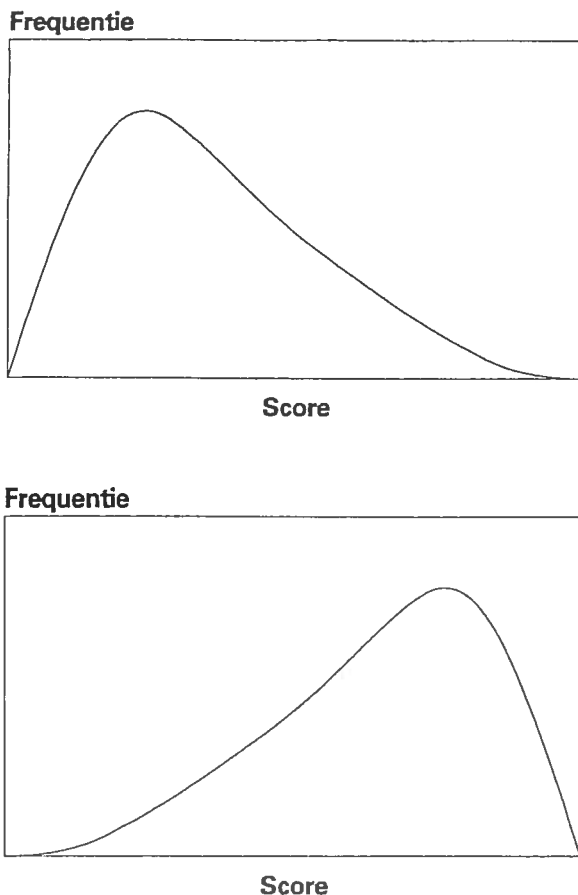
Normale en scheve verdeling

Een frequentieverdeling kan ook grafisch weergegeven worden. De lengte van een grote

groep volwassenen weergegeven in de vorm van een frequentieverdeling zal de *normaalverdeling* benaderen. Dat wil zeggen dat er weinig mensen zullen zijn die ver beneden het gemiddelde scoren en eveneens weinig mensen die ver boven het gemiddelde scoren, terwijl de meeste mensen rond het gemiddelde zullen scoren. De scores zijn dus symmetrisch verdeeld met twee ongeveer gelijke uitlopers. Bovendien is de normaalverdeling klokvormig en eentoppig. Een voorbeeld van een normaalverdeling is weergegeven in figuur 2.

**Figuur 2.** Een voorbeeld van een normaalverdeling

Voor het interpreteren van gegevens is het van groot belang na te gaan hoe de scores verdeeld zijn. Het is bijvoorbeeld mogelijk dat een frequentieverdeling van de lengte van een groep volwassenen geen normaalverdeling vertoont, maar een tweekoppige verdeling waarbij zowel



Figuur 3. Boven: een voorbeeld van een rechtsscheve verdeling; Onder: een voorbeeld van een linksscheve verdeling

de lengte van 1.68 meter als van 1.76 meter het meest voorkomt. Een mogelijke verklaring voor deze twee toppen zou kunnen zijn dat deze grafiek de verdeling van de scores van de vrouwen, met een gemiddelde lengte van 1.68 meter, en de mannen, met een gemiddelde lengte van 1.76 meter, weergeeft. Het uitdrukken van de lengte in de vorm van een gemiddelde score voor de totale groep is in dit geval niet zinvol. Voor de populatie vrouwen en mannen gelden twee verschillende gemiddelde scores. Kortom, dit voorbeeld toont het belang aan van een frequentieverdeling. Naast de normale verdeling en de tweetoppige verdeling, kan er ook sprake zijn van een *rechtsscheve* of *linksscheve* verdeling. Een voorbeeld van een rechtsscheve en linksscheve verdeling is weergegeven in figuur 3. De verdeling van scores

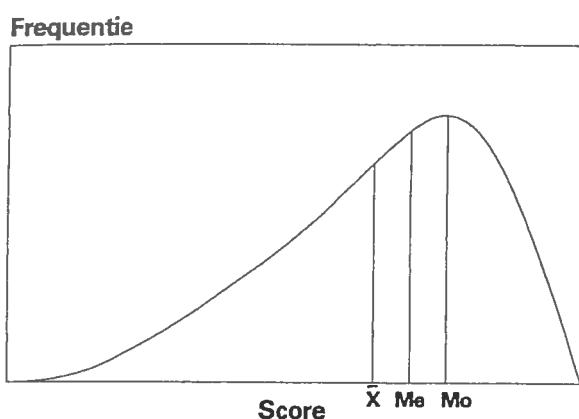
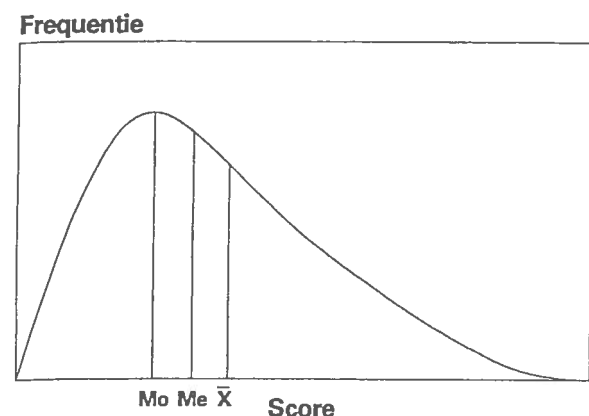
dient nagegaan te worden alvorens verdere analyses op de resultaten verricht worden, omdat een aantal statistische procedures een normale verdeling van scores veronderstelt. Voor het beoordelen of er sprake is van een normaalverdeling zijn bepaalde statistische procedures voorhanden.

Maten van centrale tendentie

Voor het weergeven van scores wordt vaak gebruik gemaakt van het *gemiddelde* (\bar{X}), dat reeds ter sprake is gekomen. Dit rekenkundig gemiddelde is een maat van centrale tendentie. Twee andere maten van centrale tendentie zijn de *modus* (Mo) en de *mediaan* (Me). De modus geeft de waarde weer die het meest voorkomt. De mediaan is de waarde waaronder en waarboven 50 procent van alle proefpersonen scoort. Hiervoor is het dus nodig de gescoorde waarden van laag naar hoog te rangschikken. In het voorbeeld in tabel 1 bedraagt het gemiddelde 6.5, de modus 7 en de mediaan ligt tussen 6 en 7. Bij een rechtsscheve verdeling is het gemiddelde hoger dan de mediaan en de mediaan hoger dan de modus. Bij een linksscheve verdeling is de gemiddelde score lager dan de mediaan en de mediaan lager dan de modus. Dit is weergegeven in figuur 4.

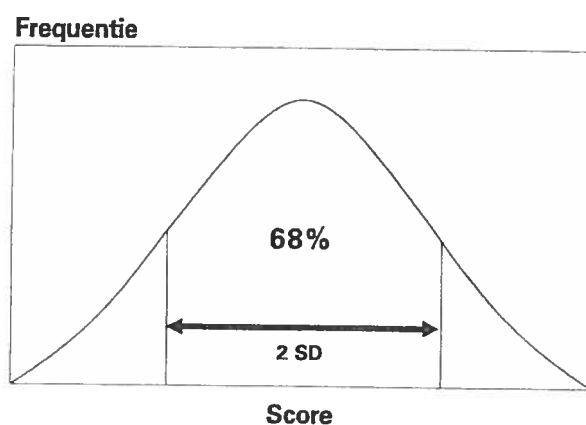
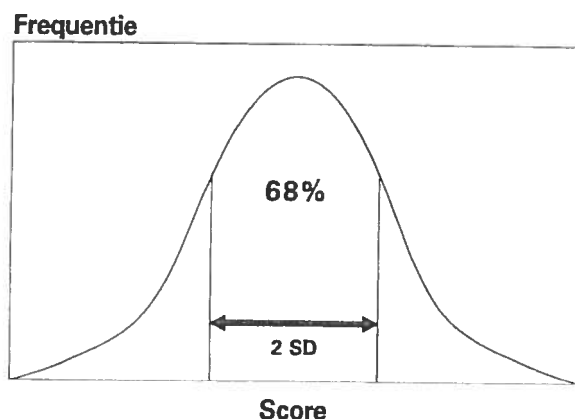
Maten van spreiding

Van belang bij het interpreteren van een gemiddelde score is de spreiding van de scores of het verschil tussen de hoogste en de laagste gemeten waarde. Dit wordt weergegeven met de *range* ofwel het scorebereik. Indien de hoogste waarde die een bepaalde variabele kan aannemen negen bedraagt en de laagste waarde drie, dan is de range zes. Een andere manier om de spreiding van scores weer te geven is de *standaarddeviatie* (SD) of standaardafwijking. De standaarddeviatie geeft de gemiddelde afwijking van de gevonden scores ten opzichte van het gemiddelde weer, uitgaande van een normale verdeling van scores. Bij een normale



Figuur 4. Boven: een voorbeeld van een rechtsscheve verdeling, waarbij het gemiddelde (\bar{X}) hoger is dan de mediaan (Me) en de mediaan hoger is dan de modus (Mo); Onder: een voorbeeld van een linksscheve verdeling, waarbij het gemiddelde (\bar{X}) lager is dan de mediaan (Me) en de mediaan lager is dan de modus (Mo)

verdeling van scores geldt dat ongeveer 68 procent van de proefpersonen tussen de gemiddelde score minus één standaarddeviatie en de gemiddelde score plus één standaarddeviatie scoort. Als de gemiddelde lengte van mannen in een bepaalde leeftijdscategorie 1.80 meter is en de standaarddeviatie 0.20 meter bedraagt, dan heeft ongeveer 68 procent van de mannen een lengte tussen de 1.60 meter en 2.00 meter. Als de lengte van een persoon binnen het bereik van één standaarddeviatie valt, dan hoort deze persoon wat betreft zijn lengte duidelijk bij de meerderheid. De grootte van de uitlopers op een normaalverdeling is van invloed op de standaarddeviatie. Zijn de uitlopers klein, dan is de standaarddeviatie klein en omgekeerd.



Figuur 5. Een voorbeeld van twee normaalverdelingen; In de bovenste grafiek is de standaarddeviatie klein en in de onderste grafiek is de standaarddeviatie groot

Bij een hoge standaarddeviatie liggen de scores ver uit elkaar, bij een lage standaarddeviatie liggen de scores dicht bij elkaar. In figuur 5 zijn twee normaalverdelingen weergegeven met een verschillende standaardafwijking. In de bovenste grafiek is de standaarddeviatie klein en in de onderste grafiek is de standaarddeviatie groot. Indien de standaarddeviatie in het beschreven voorbeeld over de lengte van mannen, niet 0.20 meter maar 0.40 meter bedraagt, dan scoort ongeveer 68 procent van de mannen tussen 1.40 en 2.20 meter. Uit dit voorbeeld blijkt dat een gemiddelde score minder eenduidig interpreteerbaar is indien de standaarddeviatie erg groot is. Tussen het gemiddelde minus twee maal de standaarddevia-

tie en het gemiddelde plus twee maal de standaarddeviatie scoort ongeveer 95 procent van de proefpersonen.

Standard Error

Uit het bovenstaande blijkt dus dat een gemiddelde score slechts een globale schatting is. Om een indruk te krijgen van de zuiverheid van deze schatting kan de *standard error* (S_e) berekend worden. De gemiddelde lengte van een groep personen noemen we het steekproefgemiddelde. Bij herhaaldelijk meten van de lengte van deze groep personen kunnen we de gevonden steekproefgemiddelden ook weer geven in de vorm van een verdeling. Deze verdeling benadert de normaalverdeling. De daarbijbehorende standaarddeviatie, dus de standaarddeviatie van de verdeling van steekproefgemiddelden, noemen we de standard error. Deze standard error geeft een indruk van de nauwkeurigheid waarmee het gemiddelde van die betreffende steekproef berekend is.

De standard error kan berekend worden door de standaarddeviatie te delen door de wortel uit de grootte van de steekproef (N). Naarmate het aantal personen in de steekproef toeneemt, wordt de meting nauwkeuriger en dus de standard error kleiner. Indien de gemiddelde score voor de steekproef gelijk is aan 80 en de standard error gelijk is aan 20, dan ligt de gemiddelde score van deze groep personen met ongeveer 68 procent zekerheid tussen 60 en 100. De gemiddelde score van deze groep personen ligt met ongeveer 95 procent zekerheid tussen 40 en 120. Het verschil tussen de standaarddeviatie en de standard error is dat de standaarddeviatie de nauwkeurigheid van de score van een individuele persoon weergeeft, terwijl de standard error de nauwkeurigheid van het gemiddelde van de steekproef weergeeft.

Opdrachten

Aan de hand van onderstaande opdrachten kan nagegaan worden in hoeverre de beschreven statistische begrippen duidelijk zijn. De antwoorden op de vragen worden weergegeven elders in dit blad. Tot slot is een aantal literatuursuggesties opgenomen.

- 1 Van welk meetniveau zijn onderstaande variabelen:
 - a Antwoordcategorieën: 1) oneens, 2) neutraal en 3) eens
 - b 3, 4, of 6 meter boven NAP
 - c Provincies
- 2 Gegeven de volgende getallenverzameling: 2, 4, 5, 6, 8, 8, 9, 9, 9, 10
 - a Wat is het gemiddelde?
 - b Wat is de mediaan?
 - c Wat is de modus?
 - d Wat voor een vorm heeft de verdeling van deze scores?
- 3 Op een toets worden gemiddeld 20 vragen goed beantwoord. De standaarddeviatie bedraagt 3 en de scores zijn normaal verdeeld.
 - a Tussen welke waarden scoort 68 procent van personen?
 - b Hoeveel procent van de personen scoort tussen de 14 en 26?

Aanbevolen Literatuur

1. Norman GR, Streiner DL. P.D.Q. Statistics. Singapore: Decker, 1986.
2. Slotboom A. Statistiek in woorden: de meest voorkomende termen en technieken. Groningen: Wolters-Noordhoff, 1987.
3. Wijvekate ML. Verklarende statistiek. Utrecht: Het Spectrum, 1982.

DE AUTEUR

D. Dolmans is als onderwijskundige verbonden aan de Faculteit der Geneeskunde van de Rijksuniversiteit Limburg.

Correspondentie-adres:

D. Dolmans, Vakgroep Onderwijsontwikkeling en -research, Postbus 616, 6200 MD Maastricht