



Beschrijvende statistiek

Eveline Wouters · Sil Aarts

Published online: 15 November 2016
© Bohn Stafleu van Loghum 2016

Samenvatting Voor het kunnen begrijpen van wetenschappelijke artikelen is het wenselijk om iets meer te weten over beschrijvende statistiek. Wat zeggen begrippen als ‘gemiddelde’, ‘mediaan’, ‘modus’, ‘interkwartielranges’ en ‘minimale en maximale waarden’ over de beschreven onderzoeksgroep en welke grafische weergaven worden daarbij gebruikt? En wat zijn ‘centrummaten’ en ‘spreidingsmaten’? In dit artikel lichten de auteurs deze begrippen toe.

We beginnen met een voorbeeld. In een (fictief) onderzoek wordt het resultaat van drie behandelingen voor voetproblematiek vergeleken: één podotherapeutische behandeling versus twee andere behandelingen (huismiddelen), behandeling A en B genoemd en weergegeven in tab. 1.

In de tabel staan de percentages, medianen, minima (min) en maxima (max), en zogenaamde IQR’s (interkwartielranges). We zullen hierna uitleggen wat

Tabel 1 Fictief onderzoek naar drie behandelingen voor voetproblematiek.

variabele	type behandeling		
	podotherapie (n = 22)	huismiddel A (n = 20)	huismiddel B (n = 21)
<i>leeftijd</i>			
mediaan	53	55	45
min–max	23–69	23–74	25–68
IQR	24,5	26,75	19
<i>geslacht</i> (% vrouw)	85	86	82
<i>BMI</i>			
mediaan	28	27	28
min–max	19–41	20–36	22–41
IQR	9,75	4,75	6,25

BMI body-mass index; *IQR* interkwartielrange (*inter quartile range*)

Voor het berekenen van centrummaten en spreidingsmaten, gaan we uit van variabelen waarmee gerekend kan worden: intervalvariabelen.

Centrummaten

Data zijn te kwantificeren met centrummaten en met spreidingsmaten. Centrummaten karakteriseren de onderzochte groep. Bijvoorbeeld voor de variabelen leeftijd en BMI in het voorbeeld. Ze geven in één keer een idee over hoe een groep er wat betreft die variabelen uitziet. In de tabel valt op dat er geen gemiddelden worden genoemd, maar medianen. Wanneer wordt een gemiddelde gegeven, wanneer een mediaan en hoe wordt een mediaan berekend?

Het gemiddelde

Om te beginnen het gemiddelde: dit wordt berekend door alle verkregen waarden bij elkaar op te tellen en te delen door het aantal verkregen waarden. Het gemiddelde is de bekendste en meest toegepaste centrummaat. Het gemiddelde wordt ook wel het ‘rekenkundig gemiddelde’ genoemd en is ook de nauwkeurigste centrummaat indien de cijfers waarover het gemiddelde wordt berekend normaal verdeeld zijn in een groep. Bij lengte betekent dat bijvoorbeeld dat er evenveel relatief grote als relatief kleine mensen in de groep zitten. Bij grote aantallen gaat dat meestal wel op: dan ligt het gemiddelde ook daadwerkelijk in het midden.

Het gemiddelde is echter niet altijd een betrouwbare maat om een groep te karakteriseren. Dat is het geval als er een aantal sterk afwijkende waarden zijn. In de praktijk is de invloed van dergelijke uitschieters, ook wel *outliers* genoemd, vooral merkbaar als het aantal gemeten waarden gering is. We geven een voorbeeld: Marieke heeft vijf toetsen gemaakt. Ze heeft de volgende cijfers behaald: 8, 9, 7, 2 en 8. Als het gemiddelde wordt berekend, zou ze uitkomen op

een 6,8. Is die 6,8 nu wel een representatieve weergave van haar prestaties? In zo’n geval bepalen we liever de mediaan.

De mediaan

Om de mediaan te bepalen, worden alle getallen eerst netjes op een rijtje gezet (in rangorde): 2-7-8-8-9. De mediaan is vervolgens het middelste getal. Het middelste getal is in dit geval een 8. NB Als het een even aantal getallen betreft, wordt de mediaan bepaald door het gemiddelde van de twee middelste getallen te nemen.

De modus

Een in de (medische) wetenschap veel minder toegepaste centrummaat, is de modus. De modus is de waarde die het meest voorkomt. Als we het voorbeeld van Marieke er nog eens bijnemen, dan is de modus in dit geval een 8. Over het algemeen is de modus geen heel betrouwbare waarneming, met als uitzondering een situatie waarin deze waarde ook heel vaak voorkomt. Een voorbeeld is het zogenaamde modaal inkomen. Het voordeel van een modus is wel dat deze voor alle soorten variabelen gebruikt kan worden, ook voor ordinale en nominale variabelen. De modus van een variabele zoals oogkleur (een nominale variabele) of opleiding (een ordinale variabele), kun je bepalen (het gemiddelde of de mediaan ervan niet). Dat is logisch, want een dergelijke variabele is niet uit te drukken in cijfers, maar er kan wel bepaald worden welke mogelijkheid het meest voorkomt.

Spreidingsmaten

Doordat centrummaten niets zeggen over uitschieters, worden ook altijd zogenaamde spreidingsmaten gegeven. Deze geven aan hoe ver de gegevens in een verdeling uit elkaar liggen. Hoe dichter gegevens bij elkaar liggen, hoe nauwkeuriger de centrummaat de uitkomsten voor de groep karakteriseert. Met andere woorden, de getallen zijn ‘stabiel’, er is weinig variatie. Als er gemiddelden worden berekend, wordt doorgaans de standaarddeviatie (SD, standaardafwijking) opgegeven als spreidingsmaat.

Standaarddeviatie

Om de standaarddeviatie, die wordt weergegeven bij het gemiddelde, te berekenen worden de volgende stappen uitgevoerd. We rekenen verder met het voorbeeld van Mariekes cijfers. Het is aardig om deze rekensom een keer te maken om er gevoel bij te krijgen. Computergestuurde statistiekprogramma’s zoals SPSS, rekenen het in de praktijk direct moeiteloos voor je uit. Bij grote aantallen scheelt dat erg veel werk.

- De afwijking van elke waarde ten opzichte van het gemiddelde wordt berekend. Mariekes gemiddelde was 6,8 en de waarden op de toetsen waren: 8, 9, 7, 2 en 8.
De berekening is dan:
 $6,8 - 8 = -1,2$; $6,8 - 9 = -2,2$; $6,8 - 7 = -0,2$; $6,8 - 8 = -1,2$; $6,8 - 2 = 4,8$.
- Deze afwijkingen worden in absolute waarden weergegeven (de 'minwaarden' worden positief gemaakt). Voorbeeld van Marieke: $1,2 + 2,2 + 0,2 + 1,2 + 4,8$.
- De afwijking in het kwadraat wordt berekend.
 $(1,2)^2 + (2,2)^2 + (0,2)^2 + (1,2)^2 + (4,8)^2 = 1,44 + 4,84 + 0,04 + 1,44 + 23,04 (= 30,80)$.
- De gemiddelde afwijking in het kwadraat wordt berekend (in dit geval het kwadraat gedeeld door het aantal toetsen dat Marieke heeft gemaakt); dit wordt de variantie genoemd:
 $30,8 : 5 = 6,16$.
- De wortel van de variantie is de standaardafwijking (SD: standaarddeviatie). De standaardafwijking is dan: $\sqrt{6,16} = 2,48$. Dus de SD voor Marieke is SD 2,48.

Opmerking We gaven al aan dat in dit geval, door de uitschieter (het cijfer 2), het gemiddelde niet zo'n goede keuze was. Dit geldt ook voor de SD. Het voorbeeld dient vooral om het principe van de berekening uit te leggen.

Interkwartielrange

Als we niet het gemiddelde, maar de mediaan als centrumwaarde nemen, dan wordt de IQR (*interquartile range*, interkwartielrange) berekend. Daarbij wordt de waarde die de hoogste 25% van alle waarden markeert (het zogenaamde 75^e percentiel) berekend, en evenzo, de waarde die de laagste 25% van alle waarden markeert (het 25^e percentiel); deze twee waarden worden van elkaar afgetrokken: dit is de IQR.

Bepalen van de mediaan en IQR van de BMI in een groep deelnemers

De gemeten waarden (BMI's) zijn: 19, 20, 24, 26, 30, 31 en 41.

De mediaan is: 26. Wat is dan de IQR?

19 - 20 - 24 - 26 - 30 - 31 - 41



20 markeert de onderste 25%,
31 de bovenste 25% (75^e percentiel):
IQR = 31 - 20 = 11.

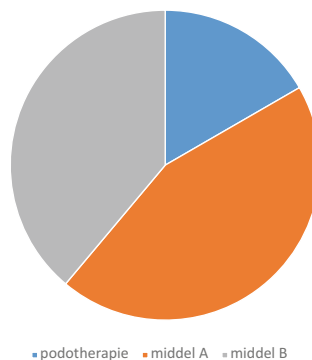
Minimale en maximale waarde

Soms wil de onderzoeker de hele range laten zien van de onderzochte populatie. Zeker als de minimale en maximale waarden niet heel ver uit elkaar liggen, geeft dat aan dat op dat punt sprake is van een homogene groep. Naast de SD en IQR als spreidingsmaten, wordt soms daarom ook de minimale en maximale waarde of de range van de uitkomsten opgegeven. In het eerste voorbeeld van het onderzoek waarbij drie behandelingen voor voetproblematiek worden gegeven, zijn de minimale en maximale waarde bijvoorbeeld voor de BMI in de podotherapiegroep 19 en 41. De 'range' (vertaald: de variatiebreedte) is hiervan direct afgeleid: het verschil tussen de maximale en minimale waarde, in het zelfde voorbeeld: $(41-19) = 22$.

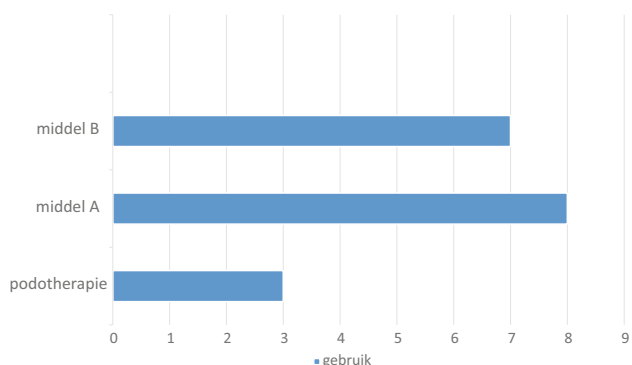
Grafieken en diagrammen

Resultaten van onderzoek worden vaak niet alleen in de tekst en tabellen, maar vaak ook in de vorm van grafieken en diagrammen weergegeven. Bekende voorbeelden zijn het histogram, het staafdiagram en het cirkeldiagram. De keuze voor welk(e) grafiek/diagram je kiest, wordt bepaald door het soort variabele waarvan de resultaten worden weergegeven.

Bij een nominale variabele, dat wil zeggen, een variabele zonder cijfermatige 'waarde', (bijvoorbeeld geslacht) wordt gebruikgemaakt van een cirkeldiagram of staafdiagram. In fig. 1 en 2 wordt van een onder-



Figuur 1 Een cirkeldiagram.



Figuur 2 Een staafdiagram.



Hier staat een advertentie.



Bohn
Stafleu
van Loghum

Springer Media

Houten 2016

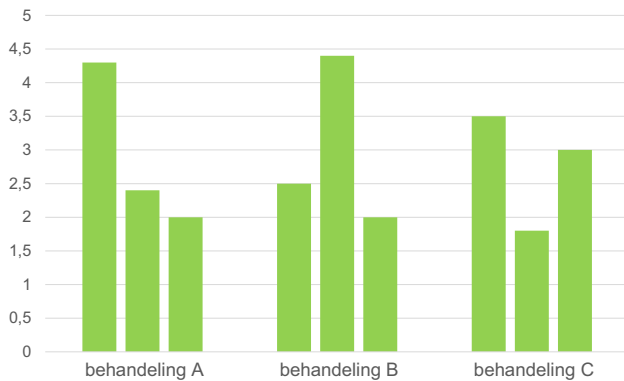
Hier staat een advertentie.



Bohn
Stafleu
van Loghum

Springer Media

Houten 2016



Figuur 3 Een histogram van pijnscores bij drie typen behandelingen.

zoek aangegeven hoe vaak patiënten voor een bepaalde voetaandoening de podotherapeut raadplegen, dan wel huismiddel A of B gebruiken respectievelijk als cirkeldiagram en als staafdiagram.

De grafische weergave van ordinale variabelen (waarvan opleidingsniveau een voorbeeld is) is ook een staafdiagram.

Ten slotte de intervalvariabelen (continue variabelen), waarbij de resultaten in cijfers kunnen worden

weergegeven (bijvoorbeeld gewicht, lengte, pijnscores). Deze kunnen worden weergegeven in een histogram. In fig. 3 wordt een fictief voorbeeld gegeven van een histogram voor de pijnscores van een populatie van patiënten met drie verschillende behandelingsvormen.

Conclusie

Bij (kwantitatief) onderzoek worden de resultaten beschreven in de vorm van centrummaten en spreidingsmaten. Afhankelijk van het aantal en type data en het voorkomen van uitschieters (outliers), wordt daarbij veelal de keuze gemaakt tussen gemiddelde of mediaan. De spreidingsmaten die daarbij horen zijn respectievelijk SD en IQR. Daarnaast wordt ook wel de minimale-maximale waarde of het verschil daartussen, de range, opgegeven. Resultaten kunnen ook grafisch worden weergegeven in de vorm van een cirkeldiagram, een staafdiagram of een histogram.

Eveline Wouters, PhDMD, arts, epidemioloog

Sil Aarts, PhD, epidemioloog, neuropsycholoog