



Gemiddelde, mediaan en spreidingsmaten

PRAKTIJKPROBLEEM

De eerste tabel van een onderzoeksartikel geeft vaak een beschrijving van de populatie waarin het onderzoek is gedaan. Voor een aantal variabelen wordt het gemiddelde en de standaarddeviatie (SD) gepresenteerd. Soms wordt geen gemiddelde gegeven maar de mediaan en wordt de SD vervangen door de range of de interkwartielafstand (interquartile range, IQR). Wat betekenen deze getallen en hoe moet je ze interpreteren?

ACHTERGROND

Het gemiddelde geeft aan wat een karakteristieke waarde is voor deze variabele in de onderzochte populatie en is de optelsom van de waarden van de afzonderlijke waarnemingen gedeeld door het aantal waarnemingen. De mediaan is de middelste waarde in de reeks. Als een deel van de uitkomsten zich aan één uiteinde van de verdeling bevindt, kan het zinvol zijn om de mediaan te rapporteren. In tegenstelling tot het gemiddelde, wordt de mediaan namelijk niet beïnvloed door deze extreme waarden.

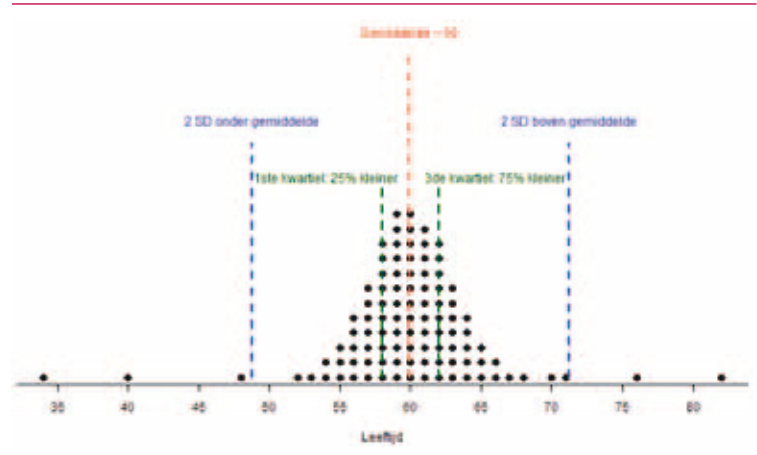
Alleen een gemiddelde of mediaan geeft geen compleet beeld van de populatie, omdat ze geen inzicht geven in hoe de waarnemingen zijn verdeeld. Daarom moet je ook kijken naar de spreidingsmaten zoals de range, de interquartile range (IQR) en de standaarddeviatie (SD). De range geeft aan wat de laagste en de hoogste waarde van de data zijn. De IQR is de afstand tussen de eerste en derde kwartiel: de grens van het eerste kwartiel is de waarde waarbij 25% van de data kleiner zijn; de grens van het derde kwartiel is de waarde waarbij 75% van de data kleiner zijn. Het nadeel van de range en de IQR is dat deze zijn gebaseerd op een beperkt aantal waarden. De SD neemt de locatie van alle waarden mee. Een grote SD betekent dat de waarnemingen gemiddeld verder van het gemiddelde afdrijven en er dus veel spreiding is. Een kleine SD houdt in dat de waarden dicht rond het gemiddelde liggen. Als de data 'normaal' zijn verdeeld, ligt 68,2% van alle waarnemingen tussen 1 SD onder en 1 SD boven het gemiddelde en ongeveer 95% van alle waarnemingen tussen 2 SD onder en 2 SD boven het gemiddelde.

UITWERKING

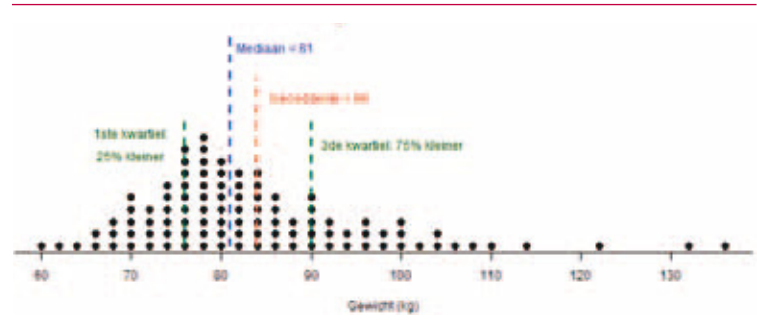
[Figuur 1] toont een populatie waarin de leeftijd normaal is verdeeld rond een gemiddelde van 60 jaar. Bij een dergelijke symmetrische verdeling is de mediaan gelijk aan het gemiddelde. De IQR loopt van 58 tot 62; 50% van de waarnemingen ligt tussen deze twee waarden. Tussen 54 (1 SD onder het gemiddelde) en 67 (1 SD boven het gemiddelde) ligt 68,2% van alle waarnemingen. In dit geval liggen de IQR en de waarden tussen -1 SD en +1 SD dicht bij elkaar. De uitkomsten van een onderzoek met deze leeftijdsverdeling van deelnemers zegt het meest over mensen tussen 54 en 67 jaar.

[Figuur 2] toont de verdeling van het gewicht in een populatie waarin deze scheef is verdeeld. Er zijn veel extreme waarden

Figuur 1 Normale verdeling



Figuur 2 Scheve verdeling



aan de rechterkant van de grafiek. In dit geval hebben het gemiddelde en de mediaan niet dezelfde waarde. Het gemiddelde (84 kg) is hoger dan de mediaan (81 kg). De IQR geeft een indruk van waar de extreme waarden liggen. De onderste IQR (76 kg) ligt dicht bij de mediaan (81 kg) dan de bovenste IQR (90 kg). Hieruit is op te maken dat het een rechtsscheve verdeling is met extreme waarden aan de rechterkant van de grafiek.

BETEKENIS

De waarden van gemiddelde, mediaan, SD, range en/of IQR in de eerste tabel van een onderzoeksartikel geven een indruk van de populatie waarin het onderzoek is verricht. Aan de hand van deze waarden kun je beoordelen of de onderzoekspopulatie overeenkomt met je eigen patiënten. ■

LITERATUUR

- 1 Swinscow TDV, Campbell MJ. Statistics at square one - Tenth edition. London: BMJ Books, 2008.

LUMC, afdeling Public Health en Eerstelijngeneeskunde, Postzone V-o-P, Postbus 9600, 2300 RC Leiden; dr. A.F.H. Smelt, huisartsdocent en epidemioloog. Vrije Universiteit Amsterdam: A.J. de Vetten, docent rekenen-wiskunde • Correspondentie: a.f.h.smelt@lumc.nl