

Hoofdstuk 3. De spreidingsmaten

Bij de centrummaten in hoofdstuk 2 ging het erom rond welke waarde op de meetschaal de waarnemingen zich concentreren of rond welke centrale waarde op de meetschaal de waarnemingen verspreid zijn. Spreidingsmaten geven aan hoe sterk de waarden zich concentreren. Spreiding is de afstand tussen de verschillende waarnemingen. Er zijn drie soorten spreidingsmaten: interkwartielafstand, varia(n)tie en standaarddeviatie. Deze maten zijn vooral van belang bij metingen op interval- en rationiveau, omdat het met name gaat om de afstand tussen de waarden.

Kwartielen

Wanneer de uitkomsten van een serie waarnemingen in oplopende volgorde worden geplaatst, ontstaat een geordende getallenreeks. Halverwege, op 50% van deze getallenreeks ligt de mediaan. Als je de getallenreeks in vier stukken verdeelt van 25% krijg je kwartielen. Kwartielen kun je niet berekenen voor nominale waarden. Het **eerste kwartiel** (Q1) is de waarde waarbij 25% bereikt wordt. Het **derde kwartiel** (Q3) is de waarde waarbij 75% bereikt wordt. Het verschil tussen het eerste en het derde kwartiel wordt de interkwartielafstand genoemd. Een **boxplot** is een grafische weergave van de kwartielen waarin de laagste waarde, het eerste kwartiel, de mediaan, het derde kwartiel en de hoogste waarde worden weergegeven.

Variatie en variantie

Men kan per onderzoekseenheid het verschil met het gemiddelde berekenen. Bijvoorbeeld voor een groep proefpersonen is de gemiddelde afstand van de tafel 2,16 meter. Persoon 1 staat 2,5 meter van de tafel af. Het verschil met het gemiddelde is dan $2,16 - 2,5 = -0,34$ meter, oftewel $x_1 - \bar{x}$. Door te kwadrateren kom je uit de - (min keer min is immers plus). Hierdoor krijg je de formule van de **variatie**: $(x_1 - \bar{x})^2$. Dit getal is als spreidingsmaat lastig te interpreteren, omdat het getal van een aantal onderzoekseenheden afhankelijk is. Hoe meer onderzoekseenheden, hoe hoger de waarde van de variatie wordt. Door de **variantie** te berekenen, wordt dit probleem opgelost. De gemiddelde kwadratische afwijking ten opzichte van het gemiddelde is de letterlijke vertaling van variantie.

De formule voor variantie:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Om iets te kunnen zeggen over de variantie, moet je op 2 punten meten en deze metingen met elkaar vergelijken. Bijvoorbeeld: de eerste is 3,21 en de tweede 5,29. Vervolgens bereken je ook de gemiddelden hiervan en vergelijk je die. Later op de avond is de spreiding tussen de personen in de kamer kleiner dan eerder op de avond, terwijl de gemiddelde afstand tot de tafel nauwelijks is veranderd (van 4,38 naar 4,5 meter).

JoHo Samenvatting – Beschrijvende Statistiek

Standaarddeviatie

Doordat er in de formule van de variatie en dus variantie gekwadrateerd wordt, is het lastig om de getallen te interpreteren. Het is daarom handig om de wortel te trekken uit de variantie. Je krijgt daardoor de gemiddelde afwijking ten opzicht van het gemiddelde: de standaarddeviatie. Hoe lager de standaarddeviatie, hoe dichter de individuele scores zich rondom het gemiddelde concentreren. Dus andersom is het dan ook: hoe hoger de standaarddeviatie hoe verder de individuele scores van het gemiddelde af liggen.

De formule voor standaarddeviatie:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Berekenen van de variantie en de standaarddeviatie in SPSS: Analyze → Descriptive Statistics → Frequencies → Statistics → Dispersion. Zo kun je de spreiding van een gegeven meten. Bijvoorbeeld de kamerprijzen in Eindhoven.

Z-scores

Dit zijn gestandaardiseerde scores van een variabele die we kunnen uitrekenen voor elke onderzoekseenheid. Hierdoor zijn de waarden van variabelen, die een verschillende meeteenheid hebben, met elkaar te vergelijken. Z geeft aan hoeveel keer de standaarddeviatie de betreffende onderzoekseenheid afwijkt van het gemiddelde van een variabele. Door naar deze meeteenheid te kijken, kan de samenhang tussen twee variabelen worden aangegeven en zal het verband duidelijker zijn dan indien alleen naar de werkelijke waarden van de variabelen wordt gekeken. Bij een negatieve z-score is de waarde van de onderzoekseenheid voor die variabele kleiner dan het gemiddelde van de groep. Bij een positieve z-score is die waarde juist groter dan het gemiddelde van de groep.

De formule voor z-score:

$$z = \frac{(x - \bar{x})}{s}$$

Berekenen van de z-score in SPSS: Analyze → Descriptive Statistics → Descriptives. De variabelen waar je z-scores van wilt hebben, voer je hier in. Vervolgens vink je het vakje Save standardized values as variables aan.

Scheve en normale verdelingen

Naast de centrum- en spreidingsmaten is ook de scheefheid ofwel symmetrie een kenmerk van verdelingen. Bij een **normale verdeling** is er sprake van een symmetrische verdeling van waarden rondom de mediaan. De modus, de mediaan en het gemiddelde vallen hierbij samen. De extreme waarden zitten aan de zogenoemde 'staartjes' van de verdeling. Van een **scheve verdeling** is sprake als de verdeling niet symmetrisch is. Scheefheid ofwel **skewness** ontstaat op het moment dat er ten opzichte van de modus aan een kant meer extreme waarden voorkomen dan aan de andere kant. Bij een positieve waarde van de skewness is de verdeling scheef naar rechts en bij een negatieve waarde is de verdeling scheef naar links.

JoHo Samenvatting – Beschrijvende Statistiek

Daarnaast kun je kijken naar de gewelddheid van de verdeling (**kurtosis**): hoe plat of spits de verdeling is. Bij een hoge kurtosis is sprake van een spitse verdeling en dus minder spreiding en extreme waarden in de verdeling. Bij een lage kurtosis is sprake van een platte verdeling, dus meer spreiding en extreme waarden.