Spreidingsmaten: variantie en standaardafwijking

Voorbeeld 1 (Examenscores gedragswetenschappen)**.** Op het examen van gedragswetenschappen scoorden de leerlingen van klas A gemiddeld 60 op 100, dus $\bar{x} = 60$.

Scores klas A: 57, 70, 55, 67, 68, 54, 63, 58, 57, 63, 65, 68, 53, 58, 56, 64, 56, 55, 54, 59

Zijn alle leerlingen ongeveer even goed, of zitten er grote verschillen tussen? We willen nagaan hoe sterk de scores rond het gemiddelde verspreid liggen. In vier stappen bouwen we twee veelgebruikte maten voor spreiding op: de *variantie* en de *standaardafwijking*.

Stap 1: Afwijkingen ten opzichte van het gemiddelde

We berekenen voor de score van elke leerling de afwijking van het gemiddelde. Dit doen we door van de score x_i van leerling met klasnummer i het gemiddelde $\bar{x}$ af te trekken:

$$x_i - \bar{x}.$$

Dit vertelt ons hoeveel punten een leerling boven of onder het gemiddelde scoorde. De eerste leerling scoorde 57. Als we daar het gemiddelde $\bar{x} = 60$ van aftrekken, krijgen we $57 - 60 = -3$.

Deze berekening maken we voor alle leerlingen.

Leerling i	1	2	3	4	5	6	7	8	9	⋯	20
Score x_i	57	70	55	67	68	54	63	58	57	⋯	59
$x_i - \bar{x}$	-3	10	-5	7	8	-6	3	-2	-3	⋯	-1

Stap 2: Negatieve tekens wegwerken

Om de totale spreiding ten opzichte van het gemiddelde te bepalen, lijkt het logisch om alle afwijkingen op te tellen. Dit heeft echter geen zin. Als we alle afwijkingen optellen, krijgen we altijd 0 omdat de positieve en negatieve afwijkingen elkaar opheffen. Dat helpt ons dus niet om de spreiding te meten.

Om dat probleem op te lossen, nemen we het kwadraat van elke afwijking.

Leerling i	1	2	3	4	5	6	7	8	9	⋯	20
Score x_i	57	70	55	67	68	54	63	58	57	⋯	59
$x_i - \bar{x}$	-3	10	-5	7	8	-6	3	-2	-3	⋯	-1
$(x_i - \bar{x})^2$	9	100	25	49	64	36	9	4	9	⋯	1

Nu zijn alle waarden positief en wegen grote afwijkingen zwaarder door.

Stap 3: Gemiddelde van de gekwadrateerde afwijkingen nemen

We nemen nu het gemiddelde van alle gekwadrateerde afwijkingen door hun som te nemen en te delen door het aantal waarnemingen. Er zitten 20 leerlingen in de klas, dus hier delen we door 20.

$$\frac{1}{20}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{20}(9 + 100 + 25 + 49 + 64 + 36 + 9 + 4 + 9 + \ldots + 1) = 28.3$$

Op deze manier hebben we de *variantie* berekend.

Definitie 1. De **variantie** meet hoe sterk de waarden afwijken van het gemiddelde. Voor n waarden $x_1, x_2, \ldots, x_n$ met gemiddelde $\bar{x}$ bereken je de variantie door:

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

CC BY-NC-SA 4.0

Stap 4: Terug naar de oorspronkelijke eenheid

De variantie heeft als eenheid het *kwadraat* van de oorspronkelijke eenheid. In ons voorbeeld is de eenheid van de variantie dus punten^2. Dat is niet zo intuïtief. We nemen de vierkantswortel om terug te keren naar de oorspronkelijke eenheid. We hebben dan de *standaardafwijking* berekend.

Definitie 2. De **standaardafwijking** is de vierkantswortel van de variantie:

$$\text{Standaardafwijking} = \sqrt{\text{Variantie}}$$

In ons voorbeeld is de variantie 28.3. Om de standaardafwijking te berekenen, nemen we de vierkantswortel.

$$\text{Standaardafwijking} = \sqrt{28.3} \approx 5.32$$

Deze standaardafwijking van 5.32 betekent dat de scores doorgaans ongeveer 5 punten van het gemiddelde van 60 verwijderd liggen.

Voorbeeld 2. Dataset: 2, 4, 6.

Het gemiddelde van deze dataset is $\bar{x} = 4$.

$$\text{Variantie} = \frac{1}{3}\left[(2-4)^2 + (4-4)^2 + (6-4)^2\right] = \frac{1}{3}(4+0+4) = \frac{8}{3} \approx 2.67$$

$$\text{Standaardafwijking} = \sqrt{\frac{8}{3}} \approx 1.63$$

Dit betekent dat de waarden in deze dataset rond het gemiddelde van 4 verspreid liggen met een standaardafwijking van ongeveer 1.63.

Oefening 1. In klas C scoren de leerlingen als volgt op het examen:

Scores klas C: 32, 23, 95, 45, 51, 21.5, 85, 100, 68, 79.5

1. Bereken het gemiddelde $\bar{x}$ van deze dataset. … …

We willen van deze dataset de variantie en standaardafwijking berekenen. Vul de ontbrekende getallen aan.

Leerling i	1	2	3	4	5	6	7	8	9	10
Score x_i	32	23	95	45	51	21.5	85	100	68	79.5
$x_i - \bar{x}$	-28	-37	35	A	-9	C	25	40	8	19.5
$(x_i - \bar{x})^2$	784	1369	1225	B	81	D	625	1600	64	380.25

2. A = … …

3. B = … …

4. C = … …

5. D = … …

6. Bereken de variantie van deze dataset.

Variantie = … …

7. Bereken de standaardafwijking van deze dataset. Rond af op twee decimalen.

Standaardafwijking = … …

CC BY-NC-SA 4.0

Spreidingsmaten: variantie en standaardafwijking

8. Wat vertellen het gemiddelde en standaardafwijking over de scores in klas C?

9. In klas A was de gemiddelde score ook 60, maar was de standaardafwijking 5.32. Wat betekent dit?

CC BY-NC-SA 4.0