Centrummaten: modus, mediaan en gemiddelde

Modus

Definitie 1. De **modus** van een dataset is de waarde die het vaakst voorkomt.

Opmerking 1. De modus is zinvol voor verschillende soorten data, zowel kwalitatief als kwantitatief. Bij *nominale* en *ordinale* variabelen geeft de modus aan welke categorie het populairst is. Bij *discrete kwantitatieve* data toont de modus de meest voorkomende waarde. Voor *continue data* is de modus minder bruikbaar, omdat er vaak veel verschillende waarden zijn die slechts één keer voorkomen.

Voorbeeld 1. Een onderzoeker bestudeert het studiegedrag van studenten.

De onderzoeker noteert hoeveel uren per week 9 studenten studeren:

8, 10, 10, 12, 14, 12, 10, 16, 18.

De modus is 10, want die waarde komt het vaakst voor, namelijk drie keer.

De onderzoeker vraagt ook naar de studenten hun favoriete studeerplek. De antwoorden zijn:

bibliotheek, bibliotheek, thuis, thuis, thuis, koffiezaak, bibliotheek, thuis, park.

De modus is *thuis*, want die studeerplek werd het vaakst genoemd.

Opmerking 2. Soms heeft een dataset meer dan één modus. Dit gebeurt wanneer meerdere waarden even vaak voorkomen en die frequentie het hoogst is. In dat geval spreken we van een *multimodale* verdeling.

Oefening 1.

1. Wat is de modus van de dataset: 15, 15, 16, 17, 17, 17, 18?

15	17	16	18

Mediaan

Definitie 2. De **mediaan** is de middelste waarde van een geordende dataset. Bij een even aantal waarden neem je het gemiddelde van de twee middelste waarden.

Opmerking 3. De mediaan is enkel zinvol bij soorten data die geordend kunnen worden, dus bij *ordinale* en *kwantitatieve* data.

CC BY-NC-SA 4.0

Voorbeeld 2. Geordende dataset met 5 elementen: 12, 12, 16, 17, 18. De mediaan is 16.

Geordende dataset met 6 elementen: 12, 12, 16, 17, 18, 20. De mediaan is $\frac{16+17}{2} = 16.5$.

Merk op dat de mediaan niet noodzakelijk een waarde is die voorkomt in de dataset.

Opmerking 4. Bij het bepalen van de mediaan is het belangrijk om *alle* waarden uit de dataset op te nemen in de geordende lijst, ook wanneer bepaalde waarden meerdere keren voorkomen. Deze herhalingen beïnvloeden de positie van het middelste element en dus de berekening van de mediaan.

Oefening 2. Wat is de mediaan van deze datasets? Orden eerst de gegevens.

1.

12, 9, 15, 10, 11

……

2.

14, 16, 10, 20, 12, 16

……

Kwartielen als verfijning van de mediaan

Definitie 3. Kwartielen verdelen een geordende dataset in vier gelijke delen:

- Het **eerste kwartiel** (Q_1) is de mediaan van de onderste helft van de data.
- Het **tweede kwartiel** (Q_2) is de mediaan van de volledige dataset.
- Het **derde kwartiel** (Q_3) is de mediaan van de bovenste helft van de data.

Voorbeeld 3. Deze dataset bevat 8 waarden: 1, 2, 4, 6, 10, 11, 14, 16. We bepalen de kwartielen als volgt:

- Q_2 is de mediaan van de volledige dataset. Omdat het aantal waarden even is, nemen we het gemiddelde van de middelste twee: $Q_2 = \frac{6+10}{2} = 8$.
- Voor Q_1 bekijken we de onderste helft van de data: 1, 2, 4, 6

  De mediaan hiervan is: $Q_1 = \frac{2+4}{2} = 3$.
- Voor Q_3 bekijken we de bovenste helft van de data: 10, 11, 14, 16

  De mediaan hiervan is: $Q_3 = \frac{11+14}{2} = 12.5$.

Samengevat: $Q_1 = 3$, $Q_2 = 8$, $Q_3 = 12.5$

CC BY-NC-SA 4.0

Opmerking 5. Bij een oneven aantal gegevens is de middelste waarde de mediaan Q_2. Deze waarde wordt uitgesloten bij het bepalen van de andere kwartielen. We berekenen Q_1 en Q_3 dan op basis van de waarden vóór en na deze mediaan.

GeoGebra link: https://www.geogebra.org/m/s7jmprwt

Oefening 3. Je kan de blauwe punten in de app hierboven verslepen of erop klikken en ze met de pijltjestoetsen verplaatsen.

1. Versleep de punten zodat $Q_1 = 5$, $Q_2 = 8$ en $Q_3 = 13$.

2. Versleep de punten zodat Q_2 precies tussen Q_1 en Q_3 ligt.

3. Versleep de punten zodat $Q_1 = Q_2 = Q_3$.

Oefening 4.

1. Bepaal de kwartielen van de geordende dataset: 2, 3, 4, 5, 10, 11, 12, 16, 17.

Wat zijn Q_1, Q_2 en Q_3?

- $Q_1 =$ ... ...
- $Q_2 =$ ... ...
- $Q_3 =$ ... ...

Gemiddelde

Definitie 4. Het **gemiddelde** van een dataset is de som van alle datawaarden x_i gedeeld door het aantal waarden. Het gemiddelde wordt typisch genoteerd met $\bar{x}$. In formulevorm kan je dit noteren als:

$$\text{Gemiddelde} = \bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Heb je moeite met het interpreteren van het symbool Σ? Doorloop dan de module over het sommatieteken.

Opmerking 6. Het gemiddelde is enkel zinvol bij kwantitatieve data, omdat je bij deze soort data op een betekenisvolle manier kunt rekenen met de waarden.

Voorbeeld 4. Dataset: 10, 12, 14, 16. De som van alle waarden is $10 + 12 + 14 + 16 = 52$.

Er zijn 4 waarden dus het gemiddelde is $\bar{x} = \frac{52}{4} = 13$.

Oefening 5.

1. Wat is het gemiddelde van: 3, 10, 12, 14, 16? ... ...

Alvorens onderstaande oefening te maken, kan het nuttig zijn om de verschillende soorten data op te frissen.

CC BY-NC-SA 4.0

Oefening 6. Welke van de centrummaten kan je bepalen bij de volgende datasets?

1. Dataset: favoriete muziekgenres van studenten te kiezen uit pop, rock, hiphop, klassiek en jazz.

Meerkeuze (selecteer alle correcte antwoorden):

(a) Modus

(b) Mediaan

(c) Gemiddelde

2. Dataset: aantal leerlingen per klas in een middelbare school.

Meerkeuze (selecteer alle correcte antwoorden):

(a) Modus

(b) Mediaan

(c) Gemiddelde

3. Dataset: energielabels van koelkasten: A++, A+, A, B, C, D, E, F.

Meerkeuze (selecteer alle correcte antwoorden):

(a) Modus

(b) Mediaan

(c) Gemiddelde

Oefening 7.

1. Een bevraging bij 11 studenten over hoeveel slaap ze gemiddeld per nacht kregen tijdens de examenperiode gaf deze resultaten:

6.5, 7, 5, 6, 6, 7.5, 6, 5.5, 6, 8, 5

Bereken de volgende centrummaten:

- **Modus:** … …
- **Mediaan:** … …
- **Gemiddelde:** (2 decimalen) … …

2. Er wordt nog een 12de student bevraagd. Hij denkt echter dat hij moet aangeven hoeveel uur hij per week slaapt en zegt **54 uur**. Deze waarde wordt toegevoegd aan de dataset. Welke van de centrummaten veranderen hierdoor?

Meerkeuze (selecteer alle correcte antwoorden):

(a) Modus

(b) Mediaan

(c) Gemiddelde

CC BY-NC-SA 4.0

> **Opmerking 7.** Bij datasets met uitschieters, zoals de foutieve waarde van 54 in de vorige oefening, verandert het gemiddelde sterk, maar de modus en mediaan blijven meestal stabiel.

Centrummaten vertellen je iets over het typische gedrag in een dataset maar ze zeggen niets over de variatie tussen de gegevens. Hiervoor kan je spreidingsmaten gebruiken.

CC BY-NC-SA 4.0