

# Handig met getallen 5

## Verbanden - 1.7 Boxplot

Rekenboek gecijferdheid voor de pabo



*Zonder verband laat alles los (Erasmusbrug Rotterdam)*

Auteur: Ruud Houweling

Adviezen: Arjen de Vetten

ISBN: 978 94 90681 289

**Voldoet aan de Kennisbasis wiskunde voor de pabo**

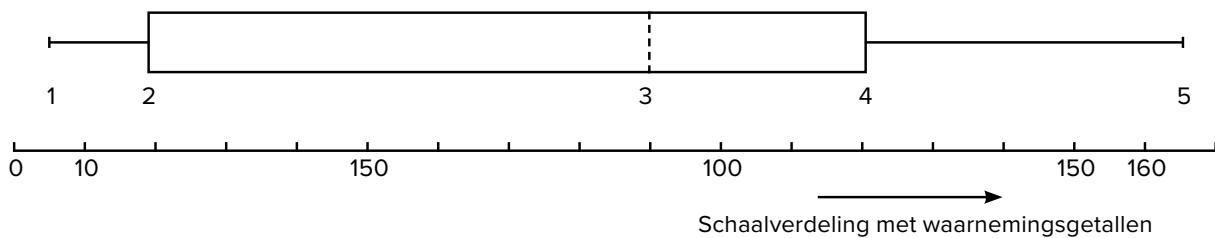
**CANTAL**  
UITGEVERIJ



## 1.7 Boxplot

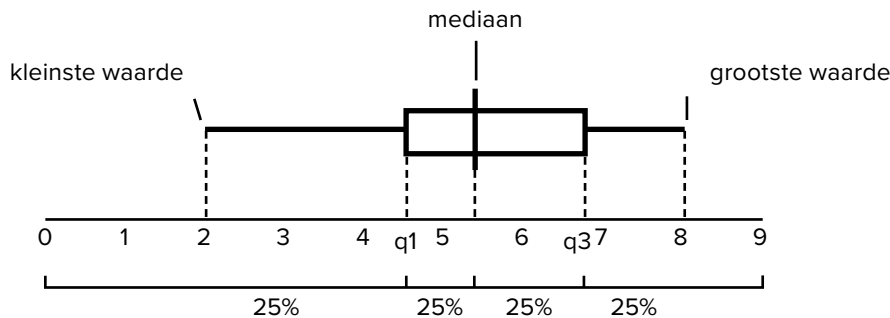
Een 'boxplot' is een grafisch model waarin je de resultaten van metingen kunt weergeven. Je kunt in een boxplot eenvoudig de laagste, hoogste en middelste (de mediaan) waarneming of meting aflezen. Bovendien krijg je een beeld van de spreiding van de waarnemingen.

Stel: je hebt door onderzoek een groot aantal waarnemingsgetallen verkregen en je wilt ze weergeven naar frequentie en spreiding. Je bepaalt dan eerst het aantal data, een andere naam voor waarnemingsgetallen. Vervolgens zet je de data op volgorde van klein naar groot en verdeel je ze in vier groepen met ieder 25% van het aantal data. De indeling heeft dus niets te maken met de waarde van de data, maar alleen met het aantal data. Zo'n groep van 25% noem je een kwartiel (een kwart van 100%). Grafisch ziet dat er zo uit:



Legenda:

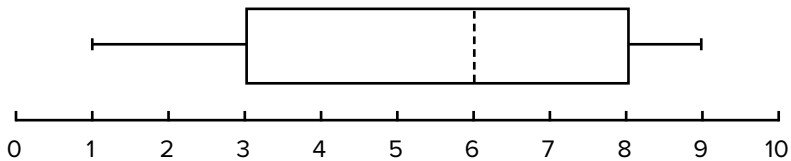
1. De kleinste waarneming (kleinste waarnemingsgetal van het onderzoek).
2. Het eerste kwartiel: de helft van de eerste helft van het aantal waarnemingen.
3. De mediaan: het middelste waarnemingsgetal.
4. Het derde kwartiel: de helft van de tweede helft van het aantal waarnemingen.
5. De grootste waarneming.



In de grafiek is de waarneming met de kleinste waarde een 2 en die met de grootste een 8. De mediaan is 5,5. Hoe je die bepaalt, heb je gelezen in de inleiding op dit hoofdstuk. Het 1e kwartiel ( $q_1$ ) loopt tot de waarde 4,5 en het 3e ( $q_3$ ) tot ongeveer 6,8.

De afbeelding laat ook zien dat de eerste 25% van de waarnemingen loopt tot  $q_1$ . Bij de mediaan is dat 50%, oftewel de helft van het aantal waarnemingen dat je in kaart brengt. Dus: 50% van de waarnemingen heeft een waarde lager dan 5,5. Bij  $q_3$  heb je 75% van alle waarnemingen weergegeven. Een boxplot geeft je ook informatie over de spreiding van de waarnemingen, dat wil zeggen: hoever de waarnemingen uit elkaar liggen. Hoe 'langer' een onderdeel van de grafiek, hoe verder de waarnemingen van dat onderdeel uit elkaar liggen.

In het voorbeeld zie je dat de lijn van de kleinste waarde tot  $q_1$  langer is die van  $q_3$  tot de grootste waarde. Dat betekent dat de spreiding van de eerste 25% van de waarnemingen groter is die van de laatste 25%. Bij de tweede 25% liggen de waarnemingen dichter bij elkaar dan bij de derde 25%. Dat zie je aan de lengte van de blokjes.

**Voorbeeld 1:** Het aflezen van een boxplot

In deze boxplot zijn de gegevens verwerkt van een statistisch onderzoek naar cijfers op de Landelijke Kennisbasis Toets (de LKT). Je ziet dat het laagste cijfer een 1 is. De mediaan lees je af bij de verticale lijn door de rechthoek: 6. De grootste waarneming, het hoogste cijfer, is een 9. De spreiding loopt van 1 tot en met 9. De ‘spreidingsbreedte’ is  $9 - 1 = 8$ .

Je kunt aan deze grafiek niet zien hoeveel keer een cijfer behaald is. Ook het gemiddelde en de modus kun je met deze gegevens niet berekenen. Wat je wel kunt aflezen is dat 25% van alle cijfers een 1, 2 of een 3 was en dat 25% van de cijfers een 3, 4, 5 of een 6 was. De volgende 25% bevatten het cijfer 6, 7 en 8 en het laatste en 4e kwartiel de cijfers 8 en 9.

Hoe kan het dat een cijfer, zoals hier o.a. de 3, in twee kwartielen voorkomt? Dat komt doordat meer dan een student een 3 heeft gehaald. Halen bijvoorbeeld 17 studenten een 3, dan kan het gebeuren, als je de resultaten op volgorde van klein naar groot zet, dat een aantal van die 17 scores in de eerste groep van 25% vallen en een aantal in het 2e kwartiel. De scheiding van de twee eerste kwartielen ligt dan bij de ‘3’. Ook de 6 en de 8 zijn in dit voorbeeld ‘grenscijfers’.

De grafiek geeft je informatie over de spreiding van de data. De onderdelen van de boxplot zijn namelijk niet allemaal even breed. Het tweede vak (van 3 tot 6) is driemaal zo breed als het laatste vak (van 8 tot 9). Conclusie: de waarnemingen in de tweede 25% liggen verder uit elkaar dan die in de laatste 25%. De middelste 50% bevat cijfers van minstens 3 tot hoogstens 8.

**Voorbeeld 2:** Het tekenen van een boxplot

Stel: Je hebt de beschikking over 54 cijfers (zonder namen) van de laatst gehouden LKT. Het zijn de cijfers (waarnemingsgetallen):

3, 5, 5, 8, 8, 4, 4, 6, 6, 5, 6, 7, 3, 5, 5, 6, 5, 7, 5, 6, 6, 9, 6, 8, 5, 7, 6, 6, 4, 8, 8, 6, 5, 7, 7, 7, 9, 5, 5, 6, 7, 6, 7, 6, 6, 8, 4, 4, 5, 5, 6, 9, 6, 5.

Je wilt er een boxplot van maken en hebt daarom deze vijf waarden nodig:

- de kleinste en de grootste waarneming;
- de mediaan;
- de waarden voor het eerste en derde kwartiel.

Eerst zet je de 54 cijfers op een rij van klein naar groot:

3, 3, 4, 4, 4, 4, 4, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9.

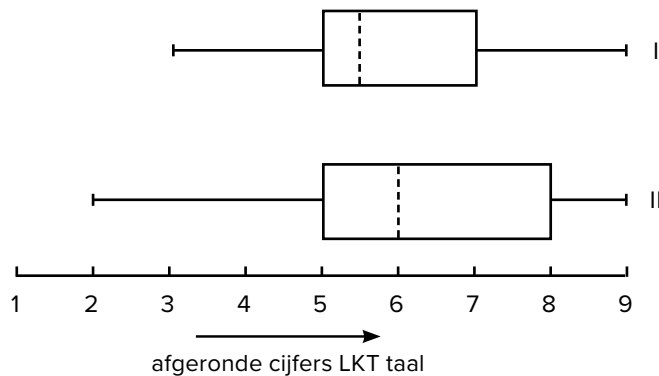
Het kleinste getal is een 3, het grootste is een 9. De mediaan is de middelste waarneming. Bij 54 waarnemingen zijn er echter twee middelste waarnemingen, namelijk de 27e en 28e. De 27e is een 6, de 28e is ook een 6. Het gemiddelde en daarmee de mediaan is dus 6.

Nu bepaal je de waardes van de middelste waarnemingen van de twee helften. De 54 waarnemingen kun je splitsen in twee helften van 27: de waarnemingen (cijfers) 1 tot en met 27 en 28 tot en met 54. Het middelste waarnemingsgetal van de eerste 27 is 14, dat is de zevende 5. Het middelste waarnemingsgetal van de tweede is 41, dat is de vijfde 7. Nu heb je alle gegevens om de boxplot te tekenen.

Teken een getallenlijn van 0 tot 10, plaats bij 3 het eerste streepje en bij 9 het laatste streepje. Het rechthoekige gebied loopt van 5 tot 7 en in het midden daarvan zet je een streepje (of stippellijn) bij de 6. Teken de boxplot op ruitjespapier en klaar!

### ► Opgave 1 Resultaten LKT

Deze tekening geeft twee boxplots weer van de uitslagen van twee afnames van de Landelijke Kennisbasis Toets taal.



Legenda:

I : resultaten afname januari

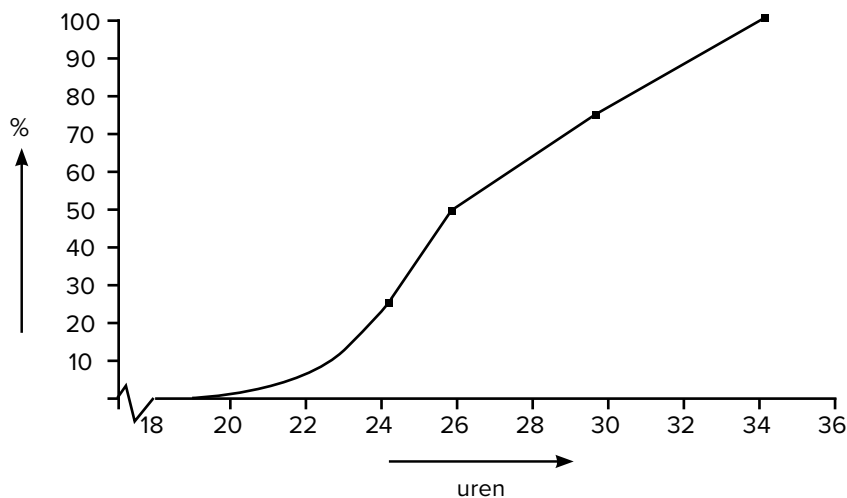
II : resultaten afname mei.

Lees de mediaan en de beide kwartielen van I af.

- Het linker lijntje van II is langer dan het rechter lijntje van II. Wat betekent dat?
- Hoe kun je zien dat er in januari een even aantal waarnemingen is verzameld?
- Hoeveel procent van de deelnemers scoorde in januari een voldoende?
- In mei was het laagste cijfer een 2. Kun je uit de boxplot afleiden hoeveel maal er een 2 is behaald? Licht je antwoord toe.

### ► Opgave 2 Eendagsvliegjes

Vorig voorjaar heeft een groot laboratoriumonderzoekplaats gevonden naar de levensduur van 'Eendagsvliegjes'. Je denkt misschien dat ze na 24 uur allemaal dood zijn, maar dat blijkt niet helemaal waar. Eendagsvliegen blijken echte 'diehards'!



In de grafiek staat op de horizontale as de levensduur, uitgedrukt in aantal uren. Op de verticale as staat het percentage vliegjes. Zo kun je bijvoorbeeld zien dat na 24 uur slechts 20% van de vliegjes dood is en dat alle vliegjes na 34 uur dood zijn. Haal de gegevens die je nodig hebt uit de grafiek en teken de boxplot van dit onderzoek op ruitjespapier.