

# Hoofdstuk 5

## Conclusies uit data

### In dit hoofdstuk

	Voorkennis	142
<b>5-1</b>	Horizontaal en verticaal percenteren	144
<b>5-2</b>	Kwantitatieve variabelen	146
<b>5-3</b>	Kwantitatieve variabelen vergelijken	148
<b>5-4</b>	Kwalitatieve variabelen	150
<b>5-5</b>	Kwalitatieve variabelen vergelijken	152
<b>5-6</b>	Data analyse	154
<b>5-7</b>	Gemengde opdrachten	156
	Samenvatting	158
	Test jezelf	160
	Extra oefening - Basis	162
	Extra oefening - Gemengd	163
	Intelligentiequotiënt	164



## Voorkennis

### Theorie

**Statistiek** is de naam van het vakgebied dat zich bezighoudt met het verzamelen, ordenen, samenvatten en analyseren van gegevens. Deze gegevens worden **data** genoemd en de verzamelde data vormen een **dataset**. Een statisticus is iemand die **statistisch onderzoek** doet. Zo'n onderzoek begint met een aantal vragen en de vaststelling over welke doelgroep het onderzoek gaat. Die doelgroep heet de **populatie**. Je gaat na op welke **statistische variabelen** het onderzoek betrekking heeft en je verzamelt data bij die variabelen. Je ordent de dataset om meer overzicht te krijgen. Uitspraken gedaan naar aanleiding van een statistisch onderzoek worden **statistische uitspraken** genoemd.

- V-1** Marloes heeft van twee kastanjabomen het gewicht van kastanjes bepaald. Boom A staat op een parkeerplaats, boom B staat in een park.
- Formuleer een mogelijke onderzoeksvraag van Marloes.
  - Wat zijn de data van haar onderzoek?
  - Op welke statistische variabelen heeft haar onderzoek betrekking?
  - Bereken de mediaan van het gewicht van de kastanjes van boom A. Bereken ook de mediaan bij boom B.
  - Het gemiddelde gewicht van de kastanjes van boom A is 6,77 gram met een standaardafwijking van 2,06. Het gemiddelde gewicht van de kastanjes van boom B is 9,16 gram met een standaardafwijking van 1,59. Welke statistische uitspraken kun je doen op grond van de data van Marloes?

gewicht in grammen	A	B
3	4	-
4	6	-
5	10	-
6	10	5
7	16	4
8	11	15
9	7	18
10	3	13
11	1	9
12	2	6
totaal	70	70

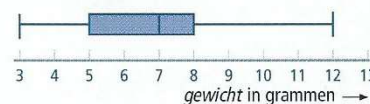
### Theorie

Een **spreidingsmaat** geeft aan of de waarnemingsgetallen dicht bij elkaar of juist ver van elkaar liggen. Spreidingsmaten zijn:

- > de **spreidingsbreedte**, dat is het verschil tussen de hoogste en de laagste waarneming;
- > de **interkwartielafstand** of **kwartielafstand**, dat is het verschil tussen het derde en het eerste kwartiel.
- > de **standaardafwijking** of **standaarddeviatie**, dat is een maat voor de spreiding ten opzichte van het gemiddelde.

- V-2a** Hiernaast zie je een boxplot bij de data van boom A van opdracht V-1.

- Hoe groot is de spreidingsbreedte? En hoe groot is de interkwartielafstand?
- Teken een boxplot bij de data van boom B.



- V-3** Gegeven zijn twee rijen getallen.
- Bereken het gemiddelde van elke rij.
  - De standaardafwijking van rij A is 3,30. Is de standaardafwijking van rij B groter of kleiner dan die van rij A? Beredeneer je antwoord.
  - Bij alle getallen in rij A wordt 2 opgeteld. Bereken het gemiddelde en de standaardafwijking van deze nieuwe rij.

<b>A</b>	10	15	15	16	17	17	22
<b>B</b>	10	10	10	16	22	22	22

**Theorie**

Frequenties kun je **absoluut** of **relatief** weergeven. De absolute frequentie is het getelde aantal, de relatieve frequentie is de verhouding van het getelde aantal tot het totaal.  
De waarneming met de grootste frequentie noem je de **modus**. Van een rij waarnemingen die op volgorde staan van klein naar groot is de middelste waarneming de **mediaan**.

**Voorbeeld**

Hieronder zie je de eindcijfers van 103 leerlingen voor het vak Frans.

<i>cijfer</i>	4	5	6	7	8
<i>frequentie</i>	9	18	30	32	14

Het absolute aantal onvoldoendes is  $9 + 18 = 27$ .

Het relatieve aantal onvoldoendes is  $\frac{27}{103} \times 100\% \approx 26\%$ .

Het gemiddelde cijfer is  $(9 \times 4 + 18 \times 5 + 30 \times 6 + 32 \times 7 + 14 \times 8) : 103 \approx 6,2$ .

De modus is het cijfer 7.

De mediaan is het 52<sup>e</sup> getal en dat is het cijfer 6.

- V-4** In de tabel hiernaast staat het aantal werklozen in de jaren 2002 en 2009, ingedeeld naar leeftijd.
- Hoe groot is de klassenbreedte?
  - Verklaar waarom de klasse 15-24 in de intervalnotatie geschreven wordt als  $[15; 25)$ .
  - Geef ook de intervalnotatie van de andere klassen.
  - Maak voor het jaar 2002 een tabel van de relatieve frequenties. Geef de relatieve frequenties in procenten, afgerond op één decimaal.
  - Maak opdracht d ook voor het jaar 2009.
  - Teken voor beide jaren een staafdiagram van de relatieve frequenties.
  - Vergelijk de staafdiagrammen met elkaar en beschrijf het verschil tussen de werkloosheid in 2002 en in 2009.

	aantal werklozen ( $\times 1000$ )	
<i>leeftijd</i>	2002	2009
15-24	79	99
25-34	74	76
35-44	74	82
45-54	56	74
55-64	19	49

- V-5** Jesper heeft het gewicht van 50 appels bepaald. De resultaten heeft hij in een steelbladdiagram verwerkt. Je ziet dat de lichtste appel 146 gram weegt.
- Hoe groot is de spreidingsbreedte?
  - En hoe groot is de interkwartielafstand?
  - Bereken de modus en de mediaan.
  - Teken een boxplot bij de gewichten van de appels.
  - Maak een klassenindeling bij het steelbladdiagram. Neem als laagste klasse  $[140; 150)$ .
  - Welke klasse is de modale klasse?

14	6 6 8 9
15	3 3 4 5 8 8 8
16	0 0 5 6 7 7 7 7 8 9 9
17	1 1 3 4 7 7 8 8 8 8 8 9 9
18	0 4 4 4 5 5 7 9 9 9
19	1 1 2 6 9