

2 De statistische cyclus

Voorkennis Centrummaten en frequentietabellen

Bladzijde 55

- 1 a Het onderzoek duurde $3 + 3 + 7 + 0 + 3 + 2 + 2 = 20$ dagen.
 b gemiddelde = $\frac{3 \cdot 0 + 3 \cdot 1 + 7 \cdot 2 + 0 \cdot 3 + 3 \cdot 4 + 2 \cdot 5 + 2 \cdot 6}{20} = 2,55$
 mediaan = $\frac{10^{\text{e}} \text{ getal} + 11^{\text{e}} \text{ getal}}{2} = \frac{2 + 2}{2} = 2$
 modus = 2

Bladzijde 56

- 2 a De modus, omdat je hier alleen kunt aflezen welke bloedgroep het meest voorkomt.
 b Bij de twee leerlingen met bloedgroep AB hoort 8%.
 Dus er zitten $\frac{2}{0,08} = 25$ leerlingen in 4H2.
 c Het verschil tussen bloedgroep A en AB is $30 - 10 = 20\%$.
 Dus er zitten $\frac{6}{0,2} = 30$ leerlingen in 4H3.
 Van deze leerlingen heeft 46,7% bloedgroep 0, dat zijn $0,467 \cdot 30 \approx 14$ leerlingen.

2.1 De onderzoeksvraag

Bladzijde 57

- 1 De logische volgorde is 3-5-6-2-4-1.

Bladzijde 58

- 2 De onderzoeksvraag: 1-3-5
 Data verzamelen: 6
 Data analyseren: 2
 Conclusies trekken: 4-1
- 3 De onderzoeksvraag: C
 Data verzamelen: D
 Data analyseren: A
 Conclusies trekken: B

Bladzijde 59

- 4 a 'hoeveel mensen' is niet scherp geformuleerd, wie bedoel je precies?
 b Dit zijn eigenlijk twee onderzoeksvragen.
 c De vraag veronderstelt dat eigenaren van elektrische auto's meer milieubewust zijn dan eigenaren van benzineauto's en dat hoeft niet zo te zijn.

Bladzijde 60

- 5 a Causaal verband. Een direct gevolg van gamen wordt onderzocht.
 b Geen causaal verband. Een samenhang wordt onderzocht, maar geen direct gevolg.
 c Causaal verband. Een direct gevolg van het verbieden van roken wordt onderzocht.
 d Geen causaal verband. Een samenhang wordt onderzocht, maar geen direct gevolg.
- 6 a Nee, want Karlijn is al langere tijd geïnteresseerd in het onderwerp en weet er al wat van af.
 b,c,d *

Bladzijde 61

- 7** Enkele voorbeelden van deelvragen:
- a** Wat valt onder verkeer? Wat is de CO₂-uitstoot van het verkeer? Wat is de totale CO₂-uitstoot?
 - b** Hoe groot is de uitstoot van fijnstof door auto's in steden? Welke gezondheidsrisico's levert fijnstof op? Bij welke hoeveelheid fijnstof treden gezondheidsrisico's op?
 - c** *
- 8** Enkele voorbeelden van deelvragen bij opgave 5:
- Onderdeel a: Hoeveel uur gamen scholieren per week? Wat wordt verstaan onder overmatig gamen? Wat wordt verstaan onder schoolprestaties?
- Onderdeel b: Hoeveel uur gamen scholieren per week? Hoeveel uur gebruiken scholieren sociale media per week?
- Onderdeel c: In hoeverre roken scholieren alleen op het schoolplein? Hoe makkelijk vinden scholieren het om ergens anders te roken? In hoeverre zijn scholieren gevoelig voor een verbod?
- Onderdeel d: Wat wordt er verstaan onder opleidingsniveaus? Hoeveel uur per week roken scholieren van verschillende opleidingsniveaus? Waar roken scholieren van verschillende opleidingsniveaus?
- 9** **a** 1-B, 2-D, 3-A, 4-C
- b** Enkele voorbeelden van deelvragen:
- 1 Wat verstaan we onder de Nederlandse woningmarkt? Hoe is de prijsontwikkeling van de Nederlandse woningmarkt van de afgelopen vijf jaren? Wat zijn de oorzaken van prijsontwikkelingen op woningmarkten?
 - 2 Wanneer is een woningmarkt stabiel? Wat zijn de effecten van verschillend beleid op woningmarkten?
 - 3 Hoe was de Nederlandse woningmarkt in 2000? Hoe is de huidige Nederlandse woningmarkt?
 - 4 Wat waren de beloftes van de regering voor de Nederlandse woningmarkt? Hoe heeft de Nederlandse woningmarkt zich ontwikkeld tijdens de regeerperiode?

Bladzijde 62

- 10** **a** Dat zijn te veel mensen, dus dat kost te veel tijd en geld.
- b** Omdat de conclusies van het onderzoek dan iets zeggen over de gehele Nederlandse bevolking.

Bladzijde 63

- 11** **a** de examenleerlingen van je school
- b** alle weekbladen
- c** de inwoners van Appelscha
- d** mensen met hooikoorts
- 12** **a** De steekproef is niet aselekt. Mensen zijn niet gelijkmatig verdeeld in leeftijd, dus in de steekproef moet daar rekening mee worden gehouden.
- b** Ja, dat ligt aan het onderzoek. De steekproef kan aselekt zijn als de populatie ook pensioengerechtigde mannen is.
- c** Ja, dat ligt aan het onderzoek. De steekproef kan groot genoeg zijn als de populatie of het aantal onderzochte kenmerken klein is.
- 13** **a** De steekproef is niet aselekt, omdat niet alle Nederlanders een even grote kans hebben om op zaterdag in de stad bij het kruispunt langs te komen.
- b** De steekproef is niet aselekt, omdat niet alle Nederlanders een even grote kans hebben in de file tijdens de ochtendspits te staan.
- c** De steekproef is te klein, waardoor niet alle meningen over klimaatverandering gegeven zullen worden.
- d** De steekproef is niet aselekt, omdat niet alle Nederlanders een even grote kans hebben om in het bezoekerscentrum van natuurgebied de Posbank te zijn.
- e** De steekproef is niet aselekt, omdat niet alle Nederlanders een even grote kans hebben om een smartphone te bezitten. Mogelijk is de steekproef ook niet groot genoeg.

- 14 Dat is meer, omdat je op grond van de peiling ongeveer $\frac{205}{1200} \times 100\% \approx 17,1\%$ VVD-stemmers verwacht.

Bladzijde 64

- 15 De populatieproportie $p = \frac{120}{10\,000} = 0,012$ en de steekproefproportie $\hat{p} = \frac{4}{250} = 0,016$.

- 16 Het aantal patiënten in het ziekenhuis is $\frac{627}{0,418} = 1500$.

Het aantal patiënten in de steekproef is $\frac{27}{0,45} = 60$.

- 17 Stel er zijn 100 personeelsleden.

Het aantal mannelijke personeelsleden met kind is dan $100 \cdot 0,3 \cdot 0,6 = 18$.

Het aantal vrouwelijke personeelsleden met kind is dan $100 \cdot 0,7 \cdot 0,5 = 35$.

Dus het aantal personeelsleden met kind is $18 + 35 = 53$.

Van het kenmerk 'personeelslid met kinderen is man' is $p = \frac{18}{53} \approx 0,34$.

Alternatieve uitwerking

Van het kenmerk 'mannelijk personeelslid heeft kinderen' is $p = 0,3 \cdot 0,6 = 0,18$.

Van het kenmerk 'vrouwelijk personeelslid heeft kinderen' is $p = 0,7 \cdot 0,5 = 0,35$.

Van het kenmerk 'personeelslid heeft kinderen' is $p = 0,18 + 0,35 = 0,53$.

Van het kenmerk 'personeelslid met kinderen is man' is $p = \frac{0,18}{0,53} \approx 0,34$.

2.2 Data verzamelen

Bladzijde 66

- 18 Bij de eerste vraag is het onduidelijk wat bedoeld wordt met 'veel'.
Bij de tweede vraag overlappen de antwoorden. Als iemand bijvoorbeeld 12 uur per week tv kijkt, dan zijn twee antwoorden op de vraag mogelijk.

Bladzijde 67

- 19 a 1. suggestieve vraagstelling
2. dubbele ontkenning
3. 'veel' is een onduidelijk begrip
4. de antwoordmogelijkheden zijn niet compleet en de vraag doet een beroep op herinnering
b *

- 20 a,b,c *

- 21 a Ja, $111 + 108 = 219$ miljoen dollar.
b Nee, de rang geeft geen hoeveelheid aan en kan niet worden opgeteld.
c eigenschappen: rang, naam en sport
hoeveelheden: inkomen (\$) en leeftijd

Bladzijde 68

- 22 a I: kwalitatief
II: kwantitatief
III: kwalitatief
IV: kwantitatief
b bijvoorbeeld: bankrekeningnummer, artikelcode, serienummer, rugnummer, pincode, zaalnummer
- 23 kwalitatief: patiëntnummer, opleiding, beroep, bloedgroep, sportfrequentie
kwantitatief: geboortjaar

Bladzijde 69

- 24** Door de laatste keuzemogelijkheid '4 of meer dagen' is de variabele kwalitatief geworden. Je kunt dan geen gemiddelde berekenen, want de antwoorden zijn geen hoeveelheden meer.
- 25** **a** Beide variabelen zijn kwalitatief.
b Bij de eerste vraag geven de antwoorden een duidelijke volgorde aan. Vaak is meer dan regelmatig, regelmatig is meer dan soms en soms is meer dan nooit. Bij de tweede vraag hebben de antwoorden geen duidelijke volgorde.
- 26** geslacht is een nominale variabele
 aantal broers/zussen is een ordinale variabele
 vooropleiding is een ordinale variabele
 woonplaats is een nominale variabele

Bladzijde 70

- 27** Enkele voorbeelden van variabelen:
 kwantitatief: reisafstand in kilometers
 nominaal: plaats van bestemming
 ordinaal: beoordeling vakantie op schaal 1-5.
- 28** De variabele 'aantal auto's' heeft alleen gehele getallen als waarden, een tussenliggende waarde is niet mogelijk. Bij de variabele 'snelheid in km/uur' zijn alle tussenliggende waarden mogelijk, een auto kan bijvoorbeeld een snelheid hebben van 85,6 km/uur.
- 29** **a** discreet
b continu
c continu
d discreet
e continu
f discreet

30 *

2.3 Centrum- en spreidingsmaten

Bladzijde 72

- 31** **a** Het meetniveau is ordinaal, want de antwoordmogelijkheden hebben een duidelijke volgorde.
b Als de helft van de mensen het helemaal eens (5) en de andere helft het helemaal oneens (1) zou zijn met een vraag, dan zijn de meningen duidelijk verdeeld. Toch zou het gemiddelde dan acceptabel (3) zijn.
c de modus

Bladzijde 73

- 32** **a** De mediaan is het meest geschikt. Het gemiddelde is gevoelig voor de uitschieter en de modus is 17, dat geeft geen goede indruk van de leeftijden.
b De modus is het meest geschikt. De andere centrummaten zijn niet te gebruiken.
c Het gemiddelde en de mediaan zijn het meest geschikt. De modus is 8, dat geeft geen goede indruk van het aantal verkochte fietsen.
d De mediaan en de modus zijn het meest geschikt. Het gemiddelde is gevoelig voor de uitschieter, die waarschijnlijk een meetfout is.
- 33** **a** cijfers op volgorde: 5,9 6,3 6,6 6,8 7,0 7,0 7,2 8,0 8,4
 mediaan = 5^e getal = 7,0
b cijfers op volgorde: 5,7 5,9 6,3 6,6 7,0 7,1 7,3 7,4
 mediaan = $\frac{4^{\text{e}} \text{ getal} + 5^{\text{e}} \text{ getal}}{2} = \frac{6,6 + 7,0}{2} = 6,8$

- c Hij kan zich het beste richten op NE, als die 6,6 een 7,1 wordt, dan heeft Nils

$$\text{mediaan} = \frac{4^{\text{e}} \text{ getal} + 5^{\text{e}} \text{ getal}}{2} = \frac{7,0 + 7,1}{2} = 7,05.$$

- 34 a de totale frequentie is $1 + 5 + 5 + 3 + 3 + 2 + 1 = 20$

$$\text{gemiddelde} = \frac{1 \cdot 0 + 5 \cdot 1 + 5 \cdot 2 + 3 \cdot 3 + 3 \cdot 4 + 2 \cdot 5 + 1 \cdot 32}{20} = 3,9$$

b $\text{mediaan} = \frac{10^{\text{e}} \text{ getal} + 11^{\text{e}} \text{ getal}}{2} = \frac{2 + 2}{2} = 2$

De leerlingen berekenen de mediaan, daaruit blijkt dat er veel minder telaarcomers zijn dan de afdelingscoördinator op basis van het gemiddelde denkt.

Bladzijde 74

- 35 a De variabele is 'aantal dagen fastfood gegeten afgelopen week'.

Omdat het een kwantitatieve variabele betreft kun je alle centrummaten berekenen.

- b totale frequentie is $2 + 4 + 6 + 5 + 4 + 4 + 3 + 2 = 30$

$$\text{gemiddelde} = \frac{2 \cdot 0 + 4 \cdot 1 + 6 \cdot 2 + 5 \cdot 3 + 4 \cdot 4 + 4 \cdot 5 + 3 \cdot 6 + 2 \cdot 7}{30} = 3,3$$

$$\text{mediaan} = \frac{15^{\text{e}} \text{ getal} + 16^{\text{e}} \text{ getal}}{2} = \frac{3 + 3}{2} = 3$$

$$\text{modus} = 2$$

op volgorde: modus, mediaan, gemiddelde

- c *

- 36 a Minstens twee keer. Hij moet twee keer het maximum van 60 punten scoren.

- b Minstens drie keer. Hij moet drie keer 55 punten of drie keer 57 punten scoren om de modus te laten veranderen, beide opties zijn goed.

- c Minstens twee keer. Hij kan met één keer spelen de mediaan maximaal verhogen naar 53 door 55 of hoger te scoren, maar het gemiddelde is dan hoger dan 53. Met twee keer spelen lukt het wel, hij kan bijvoorbeeld twee keer 54 scoren. De mediaan is dan 54 en het gemiddelde ongeveer 53,3.

- 37 a De docent heeft de mediaan berekend. De mediaan is het middelste waarnemingsgetal.

- b 25% van de leerlingen heeft een cijfer hoger dan 8,1.
50% van de leerlingen heeft een cijfer tussen 3,6 en 8,1.

Bladzijde 75

- 38 a spreidingsbreedte = $3216 - 1648 = 1568$

$$\text{interkwartielafstand} = 2304 - 1952 = 352$$

- b Dat is 50% van de werknemers, dus $0,5 \cdot 250 = 125$ werknemers.

- c Deze werknemers verdienen tussen 2304 en 3216 euro per maand.

Bladzijde 76

- 39 a Onjuist. Het is allebei 25% van de leerlingen.

- b Onjuist. Het is 25% van de leerlingen, dat zijn $0,25 \cdot 2184 = 546$ leerlingen.

- c Onjuist. Zowel binnen als buiten de box valt 50% van de leerlingen.

- d Juist. Het gemiddelde van het kleinste en grootste gewicht is $\frac{34,7 + 92,8}{2} = 63,75$ kg.

Dat valt in het laatste deel van de boxplot, dus meer dan 75% van de leerlingen weegt minder dan 63,75 kg.

- 40 a Dat is 50% van de bedrijven, dus $0,5 \cdot 18056 = 9028$ bedrijven.

- b Het gemiddelde is hoger dan de mediaan, omdat er een aantal grote bedrijven zijn tot 1250 koeien. Deze uitschieters hebben effect op het gemiddelde, maar geen effect op de mediaan.

- c Er zijn evenveel bedrijven met meer dan 117 koeien als bedrijven met minder dan 55 koeien.

- d** De bewering zou kunnen kloppen.
 Stel de bedrijven in de bovenste 25% hebben gemiddeld 600 koeien, dan hebben ze samen $0,25 \cdot 18\,056 \cdot 600 = 2\,708\,400$ koeien.
 Stel de bedrijven in de onderste 75% hebben gemiddeld 100 koeien, dan hebben ze samen $0,75 \cdot 18\,056 \cdot 100 = 1\,354\,200$ koeien.
- e** 100 ligt midden tussen 83 en 117, dus dat is ongeveer $50 + \frac{25}{2} = 62,5\%$ van de bedrijven. Een schatting van het aantal bedrijven is $0,625 \cdot 18\,056 = 11\,285$.

Bladzijde 77

- 41** In 4H-C is de minste spreiding. De meeste leerlingen scoren hier een 6 of een 7, dichtbij het gemiddelde.
 In 4H-B is de meeste spreiding. Er zijn 6 leerlingen met een 4, en 6 leerlingen met een 9, ver van het gemiddelde.
 De spreiding in 4H-A zit er tussenin.

Bladzijde 78

- 42 a** gemiddelde = $\frac{0 + 2 + 3 + 4 + 6}{5} = 3$
 De verschillen met het gemiddelde berekenen geeft -3, -1, 0, 1 en 3.
 De getallen kwadrateren geeft 9, 1, 0, 1 en 9.
 De getallen optellen geeft 20.
 Delen door het aantal geeft $\frac{20}{5} = 4$.
 De wortel van de uitkomst geeft $\sigma = \sqrt{4} = 2$.
- b** De waarnemingsgetallen in rij B zijn 10 hoger dan de waarnemingsgetallen in rij A, dus de getallen hebben dezelfde spreiding rond het gemiddelde. De standaardafwijking is de gemiddelde afwijking tot het gemiddelde en is dus voor beide rijen gelijk.
- c** Alle waarnemingsgetallen zijn gelijk aan het gemiddelde, dus $\sigma = 0$.
- 43 a** De gemiddelde afwijking van de waarnemingsgetallen tot het gemiddelde zou dan 5,3 moeten zijn. Het gemiddelde cijfer is 6,2, dus er zouden cijfers van minder dan $6,2 - 5,3 = 0,9$ en meer dan $6,2 + 5,3 = 11,5$ moeten zijn. Dat is onmogelijk.
- b** De spreiding van de waarnemingen wordt kleiner, dus de standaardafwijking wordt ook kleiner.
- 44 a** De standaardafwijking is het grootst bij groep B. De spreiding van de lengte is daar het grootst.
- b** Figuur b hoort bij Quito. De spreiding van het aantal zonuren is daar het kleinst, dus de standaardafwijking is daar ook het kleinst.

2.4 Representatie van data**Bladzijde 80**

- 45** Elk bedrag komt maar één keer voor, dus een frequentietabel geeft geen beter inzicht in de bedragen.

Bladzijde 81

- 46 a** Discrete variabele. De prijzen in een supermarkt worden op centen afgerond.
- b** in klasse $(200, 300]$
- c** De totale frequentie is $26 + 17 + 21 + 8 + 13 + 2 = 87$.
 Door klassenmiddens te gebruiken krijg je
 gemiddelde = $\frac{26 \cdot 50 + 17 \cdot 150 + 21 \cdot 250 + 8 \cdot 350 + 13 \cdot 450 + 2 \cdot 550}{87} \approx 217$ cent.
- d** De mediaan is het 44^e getal, dit ligt in de klasse $(200, 300]$.
- e** De modale klasse is $(0, 100]$.

47 a	vloeroppervlakte in m ²	frequentie
	[50, 150)	8
	[150, 250)	5
	[250, 350)	3
	[350, 450)	2
	[450, 550)	2

b De modale klasse is [50, 150).

c Het totaal aantal werknemers is $6 + 8 + 18 + \dots + 28 + 9 + 19 = 266$.

De gemiddelde vloeroppervlakte per werknemer door klassenmiddens te gebruiken is $\frac{8 \cdot 100 + 5 \cdot 200 + 3 \cdot 300 + 2 \cdot 400 + 2 \cdot 500}{266} \approx 16,9 \text{ m}^2$.

Deze waarde verschilt van het echte gemiddelde, omdat de echte waarden van de vloeroppervlakten verloren zijn gegaan bij het maken van de klassenindeling. De totale vloeroppervlakte berekend met klassenmiddens is een benadering van de werkelijke totale vloeroppervlakte.

Bladzijde 82

48 a in klasse [35, 45)

b De totale frequentie is $18 + 3 + 40 + 9 + 2 = 72$.

Door met de laagst mogelijke leeftijden binnen de klassen te rekenen krijg je het laagst mogelijke gemiddelde, dat is $\frac{18 \cdot 25 + 3 \cdot 35 + 40 \cdot 45 + 9 \cdot 55 + 2 \cdot 65}{72} \approx 41,4 \text{ jaar}$.

De gemiddelde leeftijd van de docenten is minstens 41,4 jaar, dus de bewering van de docent Frans kan niet kloppen.

c De mediaan is het gemiddelde van het 36^e en 37^e waarnemingsgetal, deze ligt in de klasse [45, 55). Dus de mediaan kan 47 jaar zijn.

d De leeftijd 36 ligt in de klasse [35, 45) en heeft daarom een maximale frequentie van 3. Binnen de klasse [45, 55) vallen maximaal 10 leeftijden (van 45 tot en met 54) met een gemiddelde frequentie van $\frac{40}{10} = 4$. Zeker één van deze leeftijden zal dus een grotere frequentie hebben dan 3. Daarom kan de bewering van de conciërge niet kloppen.

49 a modus = 18

b spreidingsbreedte = $34 - 3 = 31$

c Er zijn 30 waarnemingen.

$Q_1 = 8^{\text{e}}$ getal = 7 en $Q_3 = 23^{\text{e}}$ getal = 20
interkwartielafstand = $20 - 7 = 13$

d [0, 10), [10, 20), [20, 30) en [30, 40)

50 a 16 ouderparen

b Eén man was 33 jaar en twee vrouwen waren 35 jaar.

c zes mannen en acht vrouwen

d zes personen in de klasse [30, 36) en drie personen in de hoogste klasse [48, 52)

51 a kwantitatieve variabele

b Voor de tijden is een klassenindeling gebruikt. De getallen bij de grenzen van de staven zijn de klassengrenzen.

c De klassen sluiten op elkaar aan.

Bladzijde 84

- 52** a De variabele is gewicht in gram. Deze variabele is kwantitatief.
 b De modale klasse is $[66, 69)$.
 c het minimale gewicht is $8 \cdot 60 + 11 \cdot 63 + 13 \cdot 66 + 12 \cdot 69 + 6 \cdot 72 = 3291$ gram
 het maximale gewicht is $8 \cdot 63 + 11 \cdot 66 + 13 \cdot 69 + 12 \cdot 72 + 6 \cdot 75 = 3441$ gram
 d Bij het minimale gewicht is het gemiddelde $\frac{3291}{50} = 65,82$ gram en dat ligt buiten de klasse $[66, 69)$.
- 53** a De variabele klasse is kwalitatief, want het gaat niet om getallen.
 b Hij kan dat niet berekenen met de gegevens in de tabel, want de eerste en de laatste klasse hebben geen klassenmiddelen.
 c Omdat het om een kwalitatieve variabele gaat.
- 54** a De variabele is het gemiddeld bonbedrag in euro's. Deze variabele is continu, alle tussenliggende waarden zijn mogelijk bij een gemiddeld bedrag.
 b 10 euro, want daar begint de eerste klasse.
 c Van $6 + 7 + 4 + 3 + 5 + 3 = 28$ winkels.
 d De mediaan is het gemiddelde van het 14^e en 15^e waarnemingsgetal, dit ligt in de klasse $[30, 40)$.
 e $\frac{5}{28} \times 100\% \approx 17,9\%$

Bladzijde 85

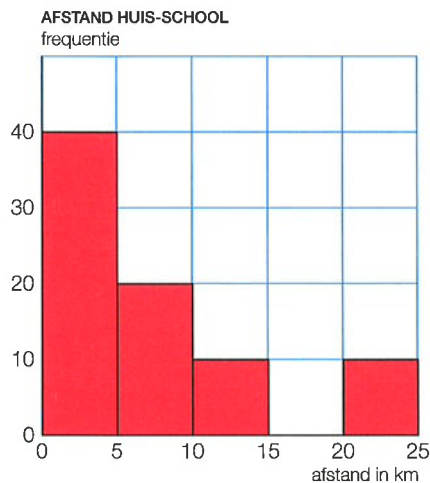
- 55** a Ja, want je hebt met een kwantitatieve variabele te maken.
 b Het totale percentage is 100.
 gemiddelde = $\frac{6 \cdot 0 + 10 \cdot 1 + 12 \cdot 2 + 14 \cdot 3 + 20 \cdot 4 + 16 \cdot 5 + 14 \cdot 6 + 8 \cdot 7}{100} = 3,76$ leerlingen
 c Er kwamen 33 dagen meer dan 5 leerlingen te laat, dus 33 dagen is $14 + 8 = 22\%$ van het onderzoek.
 Het onderzoek duurde $\frac{33}{0,22} = 150$ dagen.
 d Ze heeft $150 \cdot 3,76 = 564$ keer genoteerd.
- 56** a $0,3 + 0,7 + 2,4 + 5,9 + 12,2 = 21,5\%$ van de atleten is kleiner dan Diego.
 b $0,2 + 0,4 + 1,1 + 3,0 = 4,7\%$ van de atleten is 2 meter of langer, dus Camille hoort zelfs bij de langste 4,7% atleten.
 c $0,3 + 0,7 + 2,4 + 5,9 + 12,2 + 16,3 + 20,4 = 58,2\%$ van de atleten is kleiner dan 185 cm.
 $58,2 + 17,2 = 75,4\%$ van de atleten is kleiner dan 190.
 Dus de lengte van de atleet met percentielscore 65 valt in de klasse $[185, 190)$.
 d De mediaan heeft percentielscore 50.

Bladzijde 86

- 57** Bij het berekenen van de (relatieve) cumulatieve frequentie van een klasse tellen alle frequenties van de voorgaande klassen mee, maar ook de frequenties die binnen de klasse vallen. Door de rechterklassengrenzen te gebruiken in de (relatieve) cumulatieve frequentiepolygoon hou je daar rekening mee.
- 58** a Aflezen geeft 80 studenten.
 b Minder dan 10 km is 60 studenten.
 c Minstens 20 km is $80 - 70 = 10$ studenten.
 d Minstens 10 km en minder dan 20 km is $70 - 60 = 10$ studenten.

e

afstand in km	frequentie
$[0, 5)$	40
$[5, 10)$	20
$[10, 15)$	10
$[15, 20)$	0
$[20, 25)$	10

**Bladzijde 87**

- 59**
- a** Minder dan 100 seconden is 60%, dat zijn $0,6 \cdot 300 = 180$ klanten.
 - b** Minder dan 200 seconden is 85%, dus $0,85 \cdot 300 = 255$ klanten.
 - c** $2,5$ minuut = $2,5 \cdot 60$ seconden = 150 seconden
Minstens 150 seconden is $100 - 70 = 30\%$, dus $0,3 \cdot 300 = 90$ klanten.
 - d** Minstens 100 seconden, maar minder dan 200 seconden is $85 - 60 = 25\%$, dus $0,25 \cdot 300 = 75$ klanten.

- 60**
- a** Van 10:00 uur tot 20:00 uur is 10 uur = $10 \cdot 4$ kwartieren = 40 kwartieren.
Minder dan 20 klanten is 80%, dus $0,8 \cdot 40 = 32$ kwartieren.
 - b** Minstens 15 klanten is $100 - 40 = 60\%$, dus $0,6 \cdot 40 = 24$ kwartieren.
 - c** Minstens 10 klanten, maar minder dan 25 klanten is $90 - 20 = 70\%$, dus $0,7 \cdot 40 = 28$ kwartieren.

d

klasse	$[0, 5)$	$[5, 10)$	$[10, 15)$	$[15, 20)$	$[20, 25)$	$[25, 30)$
relatieve frequentie	10%	10%	20%	40%	10%	10%
absolute frequentie	4	4	8	16	4	4

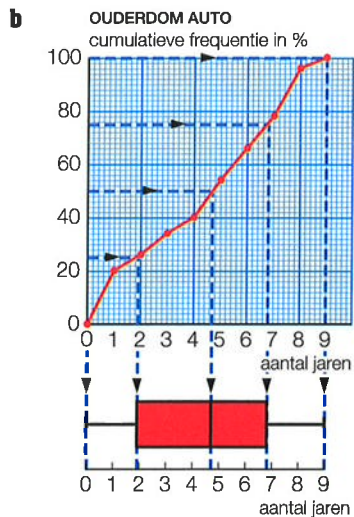
- 61** Bij Q_1 hoort 25%, bij de mediaan hoort 50% en bij Q_3 hoort 75%.

Bladzijde 88

- 62** Bij het maken van de frequentiepolygoon is een klassenindeling gebruikt, daarbij is informatie verloren gegaan. De afgelezen waarde voor de mediaan bij 50% komt uit tussen 30 en 40. De mediaan ligt dus tussen 30 en 40, maar hoeft niet precies 35 te zijn.

Bladzijde 89

- 63**
- a** Bij 25% hoort $Q_1 = 1,9$ jaar.
Bij 50% hoort mediaan = 4,7 jaar.
Bij 75% hoort $Q_3 = 6,8$ jaar.

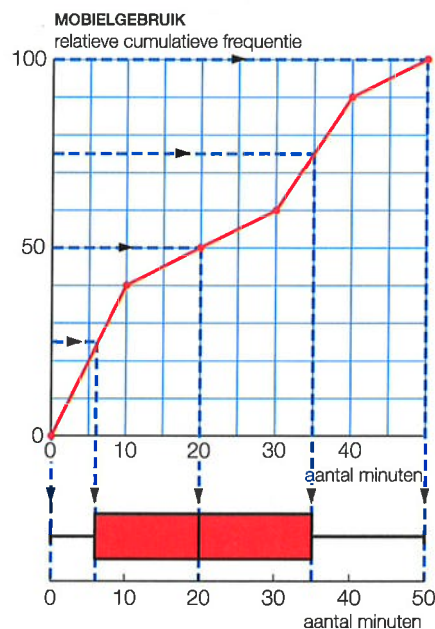


- c** Er is bij het maken van de frequentiepolygoon een klassenindeling gebruikt met een klassenbreedte van een jaar. Dan ligt het derde kwartiel ergens in de klasse $[6, 7)$, het is dan mogelijk dat een kwart van de auto's ouder is dan 6 jaar en 2 maanden.

- 64 a** Teken eerst de relatieve cumulatieve frequentiepolygoon. Bepaal daarna de waarden die nodig zijn om de boxplot te tekenen.

tijd in minuten	cumulatieve relatieve frequentie
$[0, 10)$	40
$[10, 20)$	50
$[20, 30)$	60
$[30, 40)$	90
$[40, 50)$	100

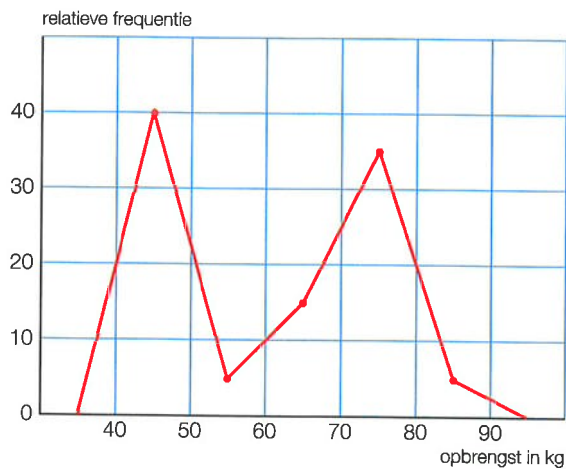
Bij 0% hoort min = 0 minuten.
 Bij 25% hoort $Q_1 = 6$ minuten.
 Bij 50% hoort mediaan = 20 minuten.
 Bij 75% hoort $Q_3 = 35$ minuten.
 Bij 100% hoort max = 50 minuten.



- b** De interkwartielafstand die uit de boxplot volgt is $35 - 6 = 29$ minuten.
c Q_1 valt in de klasse $[0, 10)$ en Q_3 valt in de klasse $[30, 40)$. De werkelijke interkwartielafstand is dus minimaal 20, maar kleiner dan 40 minuten. De werkelijke interkwartielafstand kan dus 38 minuten zijn.

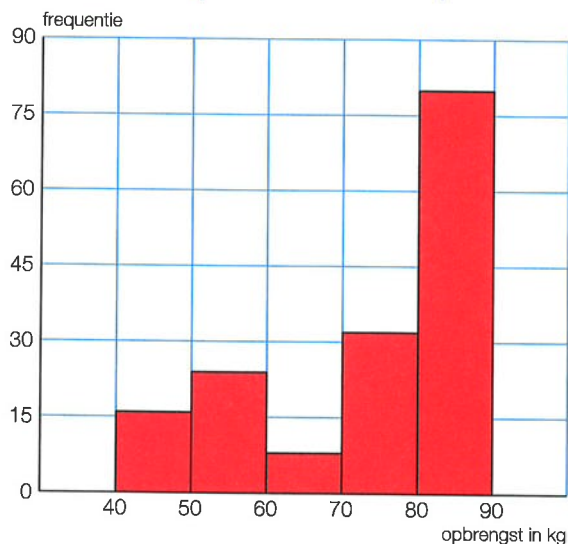
- 65 a Aanpak:** maak eerst een frequentietabel met de relatieve frequenties.

opbrengst in kg	relatieve frequentie
$[40, 50)$	40
$[50, 60)$	5
$[60, 70)$	15
$[70, 80)$	35
$[80, 90)$	5



- b De klasse $[60, 70)$ is 15%, dus $0,15 \cdot 200 = 30$ bomen.
 c *Aanpak*: maak eerst een frequentietabel met de relatieve en absolute frequenties. Bedenk dat er 160 bomen op perceel II staan.

opbrengst in kg	relatieve frequentie	absolute frequentie
$[40, 50)$	10	$0,1 \cdot 160 = 16$
$[50, 60)$	15	$0,15 \cdot 160 = 24$
$[60, 70)$	5	$0,05 \cdot 160 = 8$
$[70, 80)$	20	$0,2 \cdot 160 = 32$
$[80, 90)$	50	$0,5 \cdot 160 = 80$



d gemiddelde = $\frac{16 \cdot 45 + 24 \cdot 55 + 8 \cdot 65 + 32 \cdot 75 + 80 \cdot 85}{160} = 73,5 \text{ kg}$

2.5 Conclusies trekken

Bladzijde 91

- 66 a De variabelen bestemming en geslacht. Beide zijn kwalitatief, omdat de waarden (zeilen, Disneyland, jongen en meisje) geen hoeveelheden zijn.
 b *

Bladzijde 92

67 $\phi = \frac{21 \cdot 31 - 15 \cdot 18}{\sqrt{39 \cdot 46 \cdot 36 \cdot 49}} \approx 0,214$

Omdat $0,2 < \phi \leq 0,4$ is het verschil middelmatig.

68 a

		bijwerkingen		
		wel	niet	
genetische factor	wel	17	12	29
	niet	15	36	51
		32	48	80

$$\text{b } \phi = \frac{17 \cdot 36 - 12 \cdot 15}{\sqrt{29 \cdot 51 \cdot 32 \cdot 48}} \approx 0,287$$

Omdat $0,2 < \phi \leq 0,4$ is het verschil middelmatig.

Bladzijde 93**69**

		uitspraak		
		eens	oneens	
nationaliteit	Nederlands	620	1461	2081
	Duits	72	440	512
		692	1901	2593

$$\phi = \frac{620 \cdot 440 - 1461 \cdot 72}{\sqrt{2081 \cdot 512 \cdot 692 \cdot 1901}} \approx 0,142$$

Omdat $-0,2 \leq \phi \leq 0,2$ is het verschil gering.

70 Stel de groepen bestaan beide uit 100 personen.

		groepen		
		A	B	
mening	voor	75	25	100
	tegen	33	67	100
		100	100	200

$$\phi = \frac{75 \cdot 67 - 25 \cdot 33}{\sqrt{100 \cdot 100 \cdot 100 \cdot 100}} = 0,42$$

Omdat $\phi > 0,4$ is het verschil groot.

71 a ordinaal

b Je gebruikt de phi-coëfficiënt alleen bij twee nominale variabelen met elk twee waarden.

Bladzijde 95**72**

	Arnhem		Nijmegen		Vcp
	cum. freq.	cum. perc.	cum. freq.	cum. perc.	
zeer mee oneens	5	4%	3	2,5%	1,5%
mee oneens	11	8,8%	4	3,3%	5,5%
neutraal	23	18,4%	22	18,3%	0,1%
mee eens	61	48,8%	88	73,3%	24,5%
zeer mee eens	125	100%	120	100%	0%

$$\text{max.Vcp} = 24,5\%$$

Omdat $20 < \text{max.Vcp} \leq 40$ is het verschil middelmatig.

- 73** a ordinaal
 b Het grootste verschil in cumulatief percentage is bij één ster, dus $\max.V_{cp} = 70 - 20 = 50\%$. Omdat $\max.V_{cp} > 40$ is het verschil groot.
- 74** a De rode cumulatieve frequentiepolygoon hoort bij het kleine doelwit, omdat de boogschutters daar meer tijd nodig hadden om het doelwit te raken. Het is moeilijker om een klein doelwit te raken, dus daar doet men langer over.
 b Er zijn 200 boogschutters in elke groep. Het grootste verschil in cumulatieve frequentie is bij 5,5 seconden, dat is $150 - 30 = 120$. Dus de $\max.V_{cp} = \frac{120}{200} \times 100\% = 60\%$. Omdat $\max.V_{cp} > 40$ is het verschil groot.

Bladzijde 96

75 *

Bladzijde 97

- 76** a De boxen overlappen elkaar en geen van de medianen ligt buiten de box van de andere boxplot.
 Dus het verschil is gering.
 b De boxen overlappen elkaar en er is een mediaan die buiten de box van de andere boxplot ligt.
 Dus het verschil is middelmatig.
 c De boxen overlappen elkaar niet.
 Dus het verschil is groot.
 d De boxen overlappen elkaar en er is een mediaan die buiten de box van de andere boxplot ligt.
 Dus het verschil is middelmatig.

Bladzijde 98

- 77** a merk: kwalitatief
 inhoud: kwantitatief
 b *Aanpak*: je hebt de gemiddelden en de standaardafwijkingen, dus gebruik de effectgrootte.

$$E = \frac{1012 - 1007}{\frac{1}{2}(10 + 3)} \approx 0,769$$

Omdat $0,4 < E \leq 0,8$ is het verschil middelmatig.

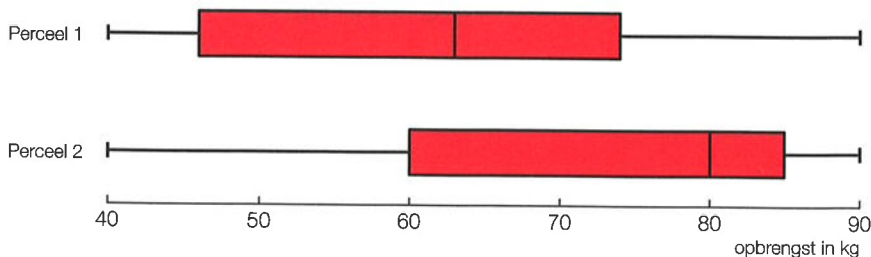
- 78** a Het grootste verschil in cumulatief percentage is bij 80 kg, dus $\max.V_{cp} = 95 - 50 = 45$. Omdat $\max.V_{cp} > 40$ is het verschil groot.

b Perceel 1

Bij 0% hoort min = 40 kg.
 Bij 25% hoort $Q_1 = 46$ kg.
 Bij 50% hoort mediaan = 63 kg.
 Bij 75% hoort $Q_3 = 74$ kg.
 Bij 100% hoort max = 90 kg.

Perceel 2

Bij 0% hoort min = 40 kg.
 Bij 25% hoort $Q_1 = 60$ kg.
 Bij 50% hoort mediaan = 80 kg.
 Bij 75% hoort $Q_3 = 85$ kg.
 Bij 100% hoort max = 90 kg.



De boxen overlappen elkaar en er is een mediaan die buiten de box van de andere boxplot ligt.
 Dus het verschil is middelmatig.

$$c \quad E = \frac{73,5 - 62,0}{\frac{1}{2}(14,2 + 14,4)} \approx 0,804$$

Omdat $E > 0,8$ is het verschil groot.

- d De effectgrootte is berekend op basis van de oorspronkelijke data en die conclusie is daarom het meest betrouwbaar.

Bij het maken van de relatieve cumulatieve frequentiepolygonen is een klassenindeling gebruikt, daarbij is informatie verloren gegaan. En bij het maken van de boxplots vanuit de relatieve cumulatieve frequentiepolygonen is nog een keer informatie verloren gegaan. Die conclusie is daarom het minst betrouwbaar.

Bladzijde 99

$$79 \quad a \quad E = \frac{3673 - 3497}{\frac{1}{2}(442 + 428)} \approx 0,405$$

Omdat $0,4 < E \leq 0,8$ is het verschil middelmatig.

- b Bij het berekenen van de effectgrootte wordt gedeeld door de som van de standaardafwijkingen. Als de standaardafwijkingen afnemen, dan deel je door een kleiner getal en wordt de effectgrootte groter.
- c De boxen overlappen elkaar en geen van de medianen ligt buiten de box van de andere boxplot.
Dus het verschil is gering.
- d Omdat E maar net groter is dan $0,4$ en de medianen ruim binnen elkaars boxen vallen is de conclusie van vraag c het best te verdedigen.

Diagnostische toets

Bladzijde 102

- 1 a de onderzoeksvraag, data verzamelen, data analyseren, conclusies trekken
- b Een causaal verband is als een gebeurtenis het directe gevolg is van een andere gebeurtenis.
Bijvoorbeeld een verband tussen de temperatuur en de verkoop van zwemkleding.
- c De steekproef moet voldoende groot en aselekt zijn.
- d De populatieproportie is het deel van de populatie dat een bepaald kenmerk heeft.
De steekproefproportie is het deel van de steekproef dat dit kenmerk heeft.
Bij een representatieve steekproef zullen deze proporties ongeveer aan elkaar gelijk zijn.

$$2 \quad a \quad p = \frac{123}{1447} \approx 0,085$$

- b Het aantal leerlingen in de steekproef is $\frac{12}{0,09} \approx 133$.

- 3 a Een kwantitatieve variabele is een hoeveelheid en een kwalitatieve variabele niet.
Een voorbeeld van een kwantitatieve variabele is lengte in cm.
Een voorbeeld van een kwalitatieve variabele is haarkleur.
- b Een voorbeeld van een nominale variabele is favoriet boek.
Een voorbeeld van een ordinale variabele is huisnummer.
- c Bij een continue variabele zijn alle tussenliggende waarden mogelijk en bij een discrete variabele niet.
Een voorbeeld van een continue variabele is voetlengte.
Een voorbeeld van een discrete variabele is schoenmaat.

- 4 a 25%, dus $0,25 \cdot 365 \approx 91$ dagen.
- b De rechtergrenzen van de kwartielen zijn $10,8^\circ\text{C}$; $15,4^\circ\text{C}$; $21,8^\circ\text{C}$ en $31,9^\circ\text{C}$.
Het gemiddelde daarvan is $\frac{10,8 + 15,4 + 21,8 + 31,9}{4} = 19,975^\circ\text{C}$.
Dus de gemiddelde temperatuur kan onmogelijk meer dan 20°C zijn geweest.
- c spreidingsbreedte $= 31,9 - -1,9 = 33,8^\circ\text{C}$ en interkwartielafstand $= 21,8 - 10,8 = 11^\circ\text{C}$
 $\frac{33,8}{11} \approx 3,07$, dus Lars heeft gelijk.
- d De beste schatting voor de standaardafwijking is 9°C .

Bladzijde 103

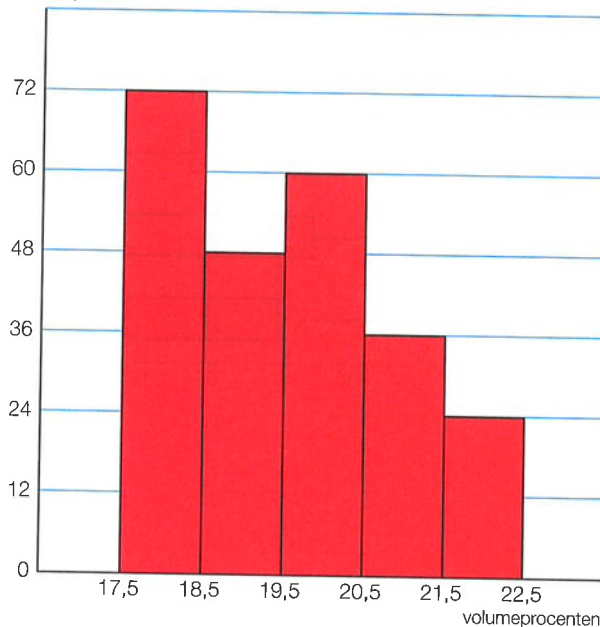
- 5 a** Het kleinst mogelijke totaal aantal klachten krijg je door te rekenen met het kleinst mogelijke aantal klachten per klasse. Dat geeft
 $2 \cdot 0 + 6 \cdot 5 + 55 \cdot 10 + 126 \cdot 15 + 135 \cdot 20 + 35 \cdot 25 + 6 \cdot 30 = 6225$ klachten.
 $6225 > 6215$ dus je kunt concluderen dat er dit jaar meer klachten waren.
- b** Er zijn 365 dagen, dus de mediaan is het 183e waarnemingsgetal;
 $2 + 6 + 55 = 63$ en $2 + 6 + 55 + 126 = 189$, dus de mediaan ligt in de klasse $[15, 20)$.
- c** De modale klasse is de klasse $[20, 25)$, want die heeft de grootste frequentie.

- 6 a** 30%, dus $0,3 \cdot 240 = 72$ mannen met minder dan 18,5%.
 25%, dus $0,25 \cdot 240 = 60$ mannen met minstens 20,5%.

b

zuurstofgehalte	relatieve frequentie	absolute frequentie
$[17,5; 18,5)$	30%	72
$[18,5; 19,5)$	20%	48
$[19,5; 20,5)$	25%	60
$[20,5; 21,5)$	15%	36
$[21,5; 22,5)$	10%	24

c ZUURSTOFGEHALTE
frequentie



7
$$\phi = \frac{34 \cdot 60 - 57 \cdot 66}{\sqrt{91 \cdot 126 \cdot 100 \cdot 117}} \approx -0,149$$

Omdat $-0,2 \leq \phi \leq 0,2$ is het verschil gering.

8
$$E = \frac{6,6 - 5,4}{\frac{1}{2}(1,2 + 1,6)} \approx 0,857$$

Omdat $E > 0,8$ is het verschil groot.

- 9 a** Bij filiaal A is de wachttijd het kortst.
b Verschil tussen A en B:
 De Vcp is maximaal bij 120 seconden.
 $\text{max.Vcp} = 90\% - 60\% = 30\%$.
 Omdat $20\% < \text{max.Vcp} \leq 40\%$ is het verschil middelmatig.

Verschil tussen B en C:
 De Vcp is maximaal bij 90 seconden.
 $\text{max.Vcp} = 50\% - 32\% = 18\%$.
 Omdat $\text{max.Vcp} \leq 20\%$ is het verschil gering.

Verschil tussen A en C:
 De Vcp is maximaal bij 120 seconden.
 $\text{max.Vcp} = 90\% - 45\% = 45\%$.
 Omdat $\text{max.Vcp} > 40\%$ is het verschil groot.

- c** Verschil tussen A en B:
- De boxen overlappen elkaar.
 - Geen van de medianen ligt buiten de box van de andere boxplot.
- Dus het verschil is gering.
 Dit is niet dezelfde conclusie als bij vraag b.

Verschil tussen B en C:

- De boxen overlappen elkaar.
- Geen van de medianen ligt buiten de box van de andere boxplot.

Dus het verschil is gering.
 Dit is dezelfde conclusie als bij vraag b.

Verschil tussen A en C:

- De boxen overlappen elkaar.
- Er is een mediaan die buiten de box van de andere boxplot ligt.

Dus het verschil is middelmatig.
 Dit is niet dezelfde conclusie als bij vraag b.

