10 Statistische variabelen

Voorkennis Soorten variabelen en de normale verdeling

Bladzijde 54

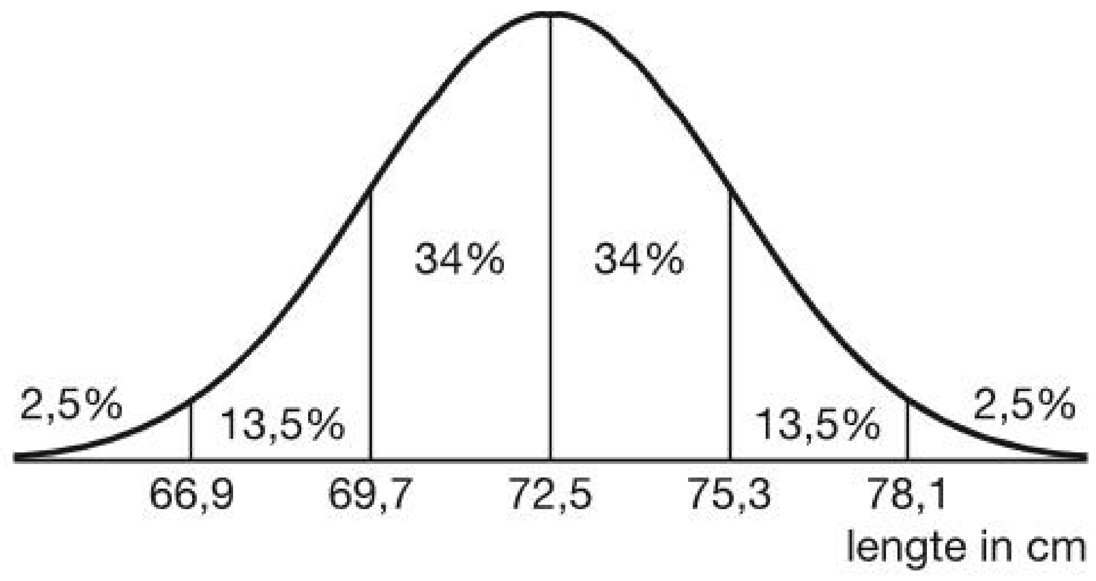
1 **a** De modus kan ook gebruikt worden bij nominale gegevens.
Voor de mediaan en het gemiddelde worden de waarden van de variabele gebruikt.
Bij de modus is dit niet het geval.

b Spreidingsbreedte, kwartielafstand en standaardafwijking.
Er is geen spreidingsmaat die je kunt gebruiken bij nominale gegevens want voor al deze spreidingsmaten is op zijn minst een rangordening in de gegevens nodig.

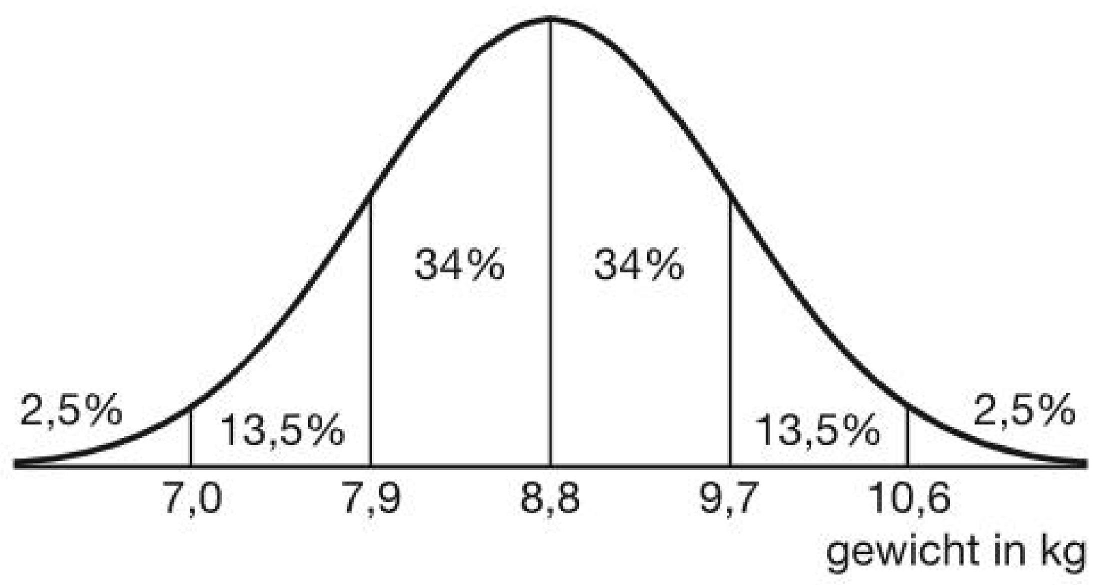
Bladzijde 55

2 **a** kwantitatief, discreet
b nominaal
c ordinaal
d kwantitatief, continu
e nominaal
f kwantitatief, discreet
g kwantitatief, continu
h ordinaal

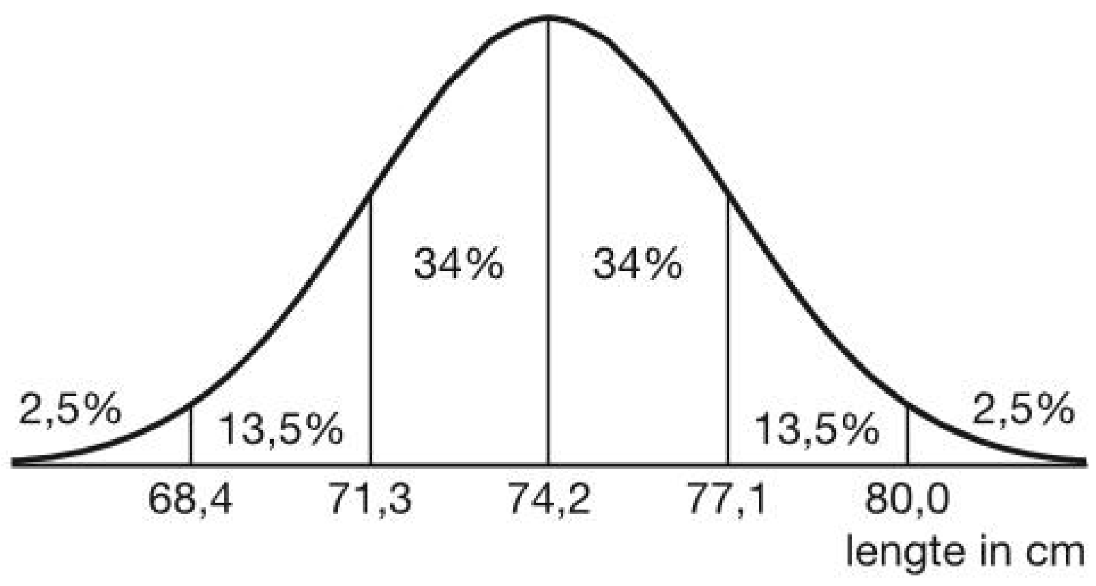
3 **a** 13,5%
b 34% + 34% + 13,5% + 2,5% = 84%



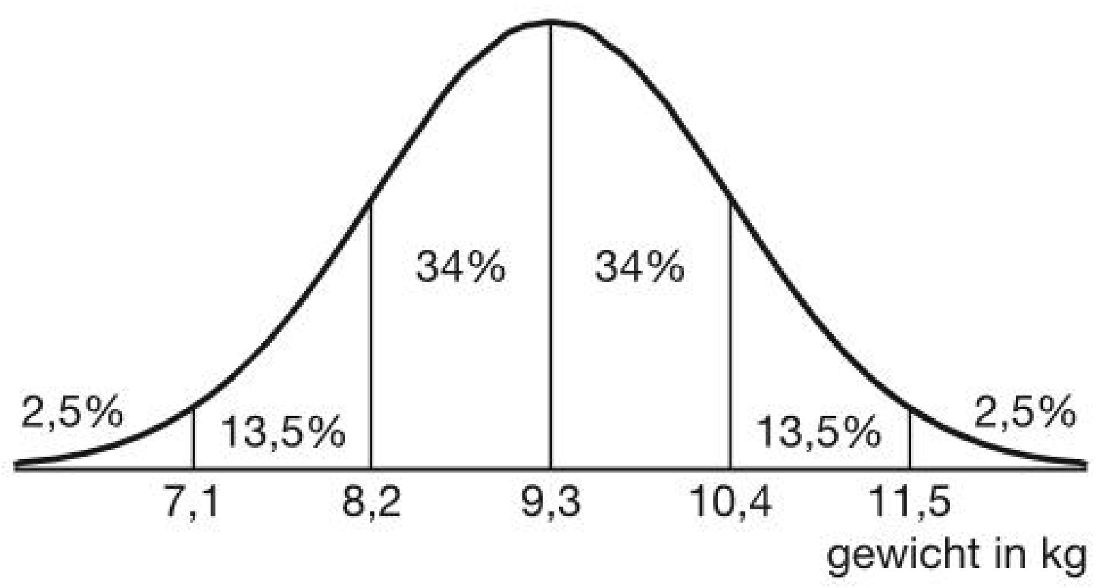
c 2,5% + 13,5% = 16%
d 2,5%



4 **a** 2,5%, dus $0{,}025 \cdot 320 = 8$ jongens
b 34% + 34% = 68%, dus $0{,}68 \cdot 320 \approx 218$ jongens



c 2,5% + 13,5% = 16%, dus $0{,}16 \cdot 320 \approx 51$ jongens
d 13,5% + 34% + 34% = 81,5%, dus $0{,}815 \cdot 320 \approx 261$ jongens



10.1 Populatie en steekproef

Bladzijde 56

1 a $\hat{p} = \frac{117}{150} = 0{,}78$

b $\sigma = \sqrt{\frac{0{,}78 \cdot 0{,}22}{150}} = 0{,}0338...$

$\hat{p} - 2\sigma = 0{,}78 - 2 \cdot 0{,}0338... \approx 0{,}712$
$\hat{p} + 2\sigma = 0{,}78 + 2 \cdot 0{,}0338... \approx 0{,}848$
Het 95%-betrouwbaarheidsinterval is [0,712; 0,848].

Bladzijde 57

2 a $\hat{p} = \frac{225}{1250} = 0{,}18$ en $\sigma = \sqrt{\frac{0{,}18 \cdot 0{,}82}{1250}} = 0{,}0108...$

$\hat{p} - 2\sigma = 0{,}18 - 2 \cdot 0{,}0108... \approx 0{,}158$
$\hat{p} + 2\sigma = 0{,}18 + 2 \cdot 0{,}0108... \approx 0{,}202$
Het 95%-betrouwbaarheidsinterval is [0,158; 0,202].

b $\hat{p} = \frac{143}{715} = 0{,}2$ en $\sigma = \sqrt{\frac{0{,}28 \cdot 0{,}8}{750}} = 0{,}0149...$

$\hat{p} - 2\sigma = 0{,}2 - 2 \cdot 0{,}0149... \approx 0{,}170$
$\hat{p} + 2\sigma = 0{,}2 + 2 \cdot 0{,}0149... \approx 0{,}230$

Het 95%-betrouwbaarheidsinterval is [0,170; 0,230].

c Van de 225 personen uit Zuid-Holland waren er 143 man en 225 − 143 = 82 vrouw.

$\hat{p} = \frac{82}{535} = 0{,}153...$ en $\sigma = \sqrt{\frac{0{,}153... \cdot (1 - 0{,}153...)}{535}} = 0{,}0155...$

$\hat{p} - 2\sigma = 0{,}153... - 2 \cdot 0{,}0155... \approx 0{,}122$
$\hat{p} + 2\sigma = 0{,}153... + 2 \cdot 0{,}0155... \approx 0{,}184$
Het 95%-betrouwbaarheidsinterval is [0,122; 0,184].

d De breedte van [0,158; 0,202] is 0,202 − 0,158 = 0,044.
De breedte van [0,170; 0,230] is 0,230 − 0,170 = 0,06.
De breedte van [0,122; 0,184] is 0,184 − 0,122 = 0,062.
Van smal naar breed: [0,158; 0,202], [0,170; 0,230], [0,122; 0,184].
Hoe groter de steekproefomvang hoe smaller het interval.

e $4\sigma = 0{,}264 - 0{,}216 = 0{,}048$, dus $\sigma = \frac{0{,}048}{4} = 0{,}012$.

f $4\sigma = 0{,}260 - 0{,}220 = 0{,}04$, dus

$\hat{p} = \frac{0{,}220 + 0{,}260}{2} = 0{,}24$

$\sigma = \sqrt{\frac{0{,}24 \cdot 0{,}76}{n}} = \sqrt{\frac{0{,}1824}{n}}$

Je krijgt de vergelijking $\sqrt{\frac{0{,}1824}{n}} = 0{,}01$.

Voer in $y_1 = \sqrt{\frac{0{,}1824}{x}}$ en $y_2 = 0{,}01$.
Intersect geeft $x = 1824$.
Men had dus 1824 festivalgangers moeten ondervragen.

g $\sigma = \sqrt{\dfrac{\hat{p} \cdot (1 - \hat{p})}{415}}$ en $\sigma = 0{,}020$

Je krijgt de vergelijking $\sqrt{\dfrac{\hat{p} \cdot (1 - \hat{p})}{415}} = 0{,}020$.

Voer in $y_1 = \sqrt{\dfrac{x \cdot (1 - x)}{415}}$ en $y_2 = 0{,}020$.
Intersect geeft $x = 0{,}21017...$ of $x = 0{,}7898...$
De oplossing $\hat{p} = 0{,}7898...$ ligt ver buiten [0,216; 0,264] en is dus zeer onwaarschijnlijk.
Dus naar verwachting hadden $0{,}21017... \cdot 415 \approx 87$ van hen een leeftijd van 18-24 jaar.

Bladzijde 58

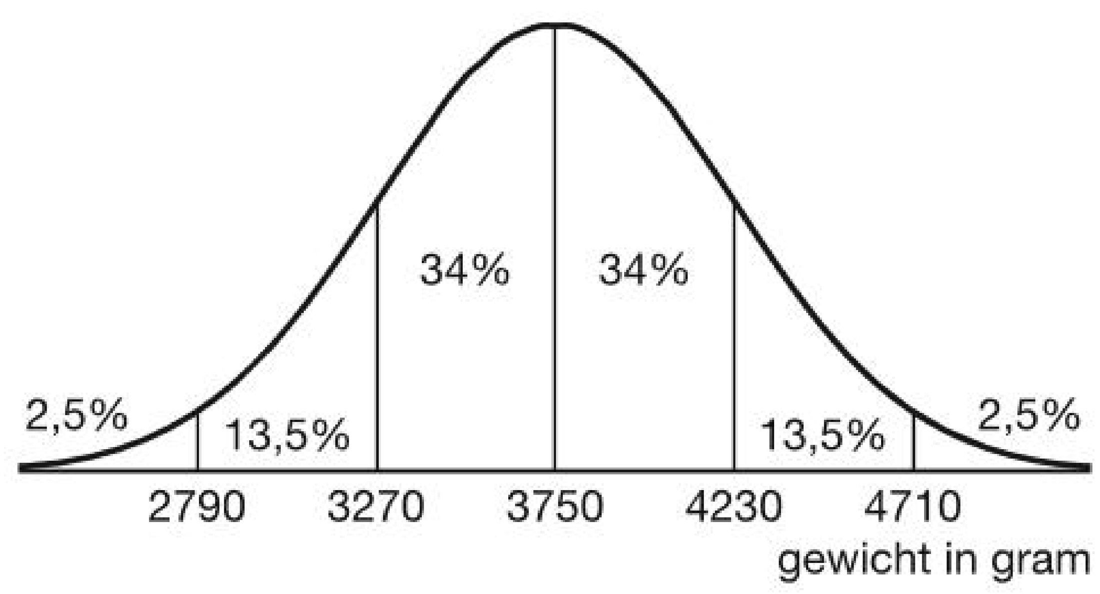
3 **a** $\hat{p} = \dfrac{104}{1908} = 0{,}0545...$ en $\sigma = \sqrt{\dfrac{0{,}0545...(1 - 0{,}0545...)}{1908}} = 0{,}0051...$

$\hat{p} - 2\sigma = 0{,}0545... - 2 \cdot 0{,}0051... \approx 0{,}0441$
$\hat{p} + 2\sigma = 0{,}0545... + 2 \cdot 0{,}0051... \approx 0{,}0649$
Het 95%-betrouwbaarheidsinterval is [0,0441; 0,0649].

b $0{,}0649 \cdot 55\,000 = 3569{,}5$, dus 3570 buitenlandse bezoekers.

c Bij de ondergrens van het 95%-betrouwbaarheidsinterval van vraag a hoort een aantal van 2426 en bij de bovengrens een aantal van 3570 buitenlandse bezoekers (zie b).
Het 95%-betrouwbaarheidsinterval voor het aantal bezoekers is dus [2426, 3570].
Omdat het 95%-betrouwbaarheidsinterval voor het aantal bezoekers ook aantallen kleiner dan 2500 en groter dan 3499 bevat kan het aantal bezoekers volgens dit interval, afgerond op duizendtallen, dus ook 2000 of 4000 zijn.

4 **a** $3750 + 2 \cdot 480 = 4710$, dus 2,5% was naar verwachting zwaarder dan 4710 gram. Dat zijn $0{,}025 \cdot 87\,957 \approx 2200$ jongens.



b $\dfrac{12\,800}{18\,852} \times 100\% \approx 68\%$.
De gewichten wijken dus hoogstens 1σ af van het gemiddelde.
$\mu - \sigma = 3750 - 480 = 3270$ gram
$\mu + \sigma = 3750 + 480 = 3230$ gram
De gewichten van deze jongens liggen tussen 3270 gram en 4230 gram.

Bladzijde 60

5 28 jaar en 4 maanden is $28 \times 12 + 4 = 340$ maanden en 3,5 jaar is 42 maanden.
Het 95%-betrouwbaarheidsinterval is $\left[340 - 2 \cdot \dfrac{42}{\sqrt{215}}, 340 - 2 \cdot \dfrac{42}{\sqrt{215}}\right] \approx [334, 346]$.
334 maanden is 27 jaar en 10 maanden.
346 maanden is 28 jaar en 10 maanden.
Met 95% zekerheid ligt de gemiddelde leeftijd waarop vrouwen voor het eerst in het huwelijk treden tussen 27 jaar en 10 maanden en 28 jaar en 10 maanden.

6 **a** $\bar{X} = \dfrac{179{,}8 + 170{,}2}{2} = 175$

$4 \cdot \dfrac{S}{\sqrt{35}} = 179{,}8 - 170{,}2$ ofwel $4 \cdot \dfrac{S}{\sqrt{35}} = 9{,}6$

Voer in $y_1 = 4 \cdot \dfrac{x}{\sqrt{35}}$ en $y_2 = 9{,}6$.
Intersect geeft $x \approx 14{,}2$.
Dus $S \approx 14{,}2$.
Het steekproefgemiddelde is 175 en de steekproefstandaardafwijking is 14,2.

b $\bar{X} = \dfrac{175 + 185}{2} = 180$

$4 \cdot \dfrac{10}{\sqrt{n}} = 185 - 175$ ofwel $4 \cdot \dfrac{10}{\sqrt{n}} = 10$.

Voer in $y_1 = 4 \cdot \dfrac{10}{\sqrt{x}}$ en $y_2 = 10$.

Intersect geeft $x = 16$.

Het steekproefgemiddelde is 180 en de steekproefomvang is 16.

7 De breedte van een 95%-betrouwbaarheidsinterval voor een gemiddelde is $4 \cdot \dfrac{S}{\sqrt{n}}$.

De breedte van het 95%-betrouwbaarheidsinterval voor meisjes van 6 maanden oud is $4 \cdot \dfrac{2{,}8}{\sqrt{38}} \approx 1{,}817$.

De breedte van het 95%-betrouwbaarheidsinterval voor meisjes van 12 maanden oud is $4 \cdot \dfrac{3{,}2}{\sqrt{38}} \approx 2{,}076$.

De breedte van het 95%-betrouwbaarheidsinterval voor jongens van 6 maanden oud is $4 \cdot \dfrac{3{,}1}{\sqrt{43}} \approx 1{,}891$.

De breedte van het 95%-betrouwbaarheidsinterval voor jongens van 12 maanden oud is $4 \cdot \dfrac{3{,}4}{\sqrt{43}} \approx 2{,}074$.

Het interval voor meisjes van 6 maanden oud is het smalst.

10.2 Groepen vergelijken

Bladzijde 62

8 **a** geslacht en bestemming

b *

Bladzijde 64

9 **a** Met de regel ‘delen door een breuk is vermenigvuldigen met het omgekeerde van die breuk’ krijg je

$OR = \dfrac{\frac{a}{b}}{\frac{c}{d}} = \dfrac{a}{b} \cdot \dfrac{d}{c} = \dfrac{ad}{bc}$ en ad en bc zijn de kruisproducten.

10

b $PV = \dfrac{485}{3707} \times 100\% - \dfrac{180}{6293} \times 100\% \approx 10{,}2\%$

$15\% < PV \leq 30\%$ en dit wijst op een middelmatig verschil.

Dus het percentageverschil geeft dezelfde conclusie over het verschil.

10 **a** Van de Nederlanders is $\dfrac{620}{2081} \times 100\% \approx 29{,}8\%$ het met de uitspraak eens.

Van de Duitsers is dat $\dfrac{72}{512} \times 100\% \approx 14{,}1\%$.

Het percentageverschil is $29{,}8\% - 14{,}1\% = 15{,}7\%$.

Omdat $15\% < PV \leq 30\%$ is het verschil middelmatig.

b $OR = \dfrac{620 \cdot 440}{1461 \cdot 72} \approx 2{,}6$

Omdat $2 \leq OR \leq 3$ is het verschil middelmatig.

$phi = \dfrac{620 \cdot 440 - 1461 \cdot 72}{\sqrt{2081 \cdot 512 \cdot 692 \cdot 1901}} \approx 0{,}14$

Omdat $-0{,}2 < phi < 0{,}2$ is het verschil gering.

Bij gebruik van de OR trek je dezelfde conclusie als bij vraag a maar bij gebruik van de phi-coëfficiënt niet.

Bladzijde 65

11 **a**

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

b $PV = \frac{17}{29} \times 100\% - \frac{15}{51} \times 100\% \approx 29{,}2\%$

Omdat $15\% < PV \leq 30\%$ is het verschil middelmatig.

$OR = \frac{17 \cdot 36}{12 \cdot 15} = 3{,}4$

Omdat $OR > 3$ is het verschil groot.

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

Omdat $0{,}2 < phi < 0{,}4$ is het verschil middelmatig.

Het verschil is middelmatig.

12 **a** ordinaal

b *

Bladzijde 69

13 **a** • De boxen overlappen elkaar.
• Geen van de medianen ligt buiten de box van de andere boxplot.
Dus het verschil is gering.

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

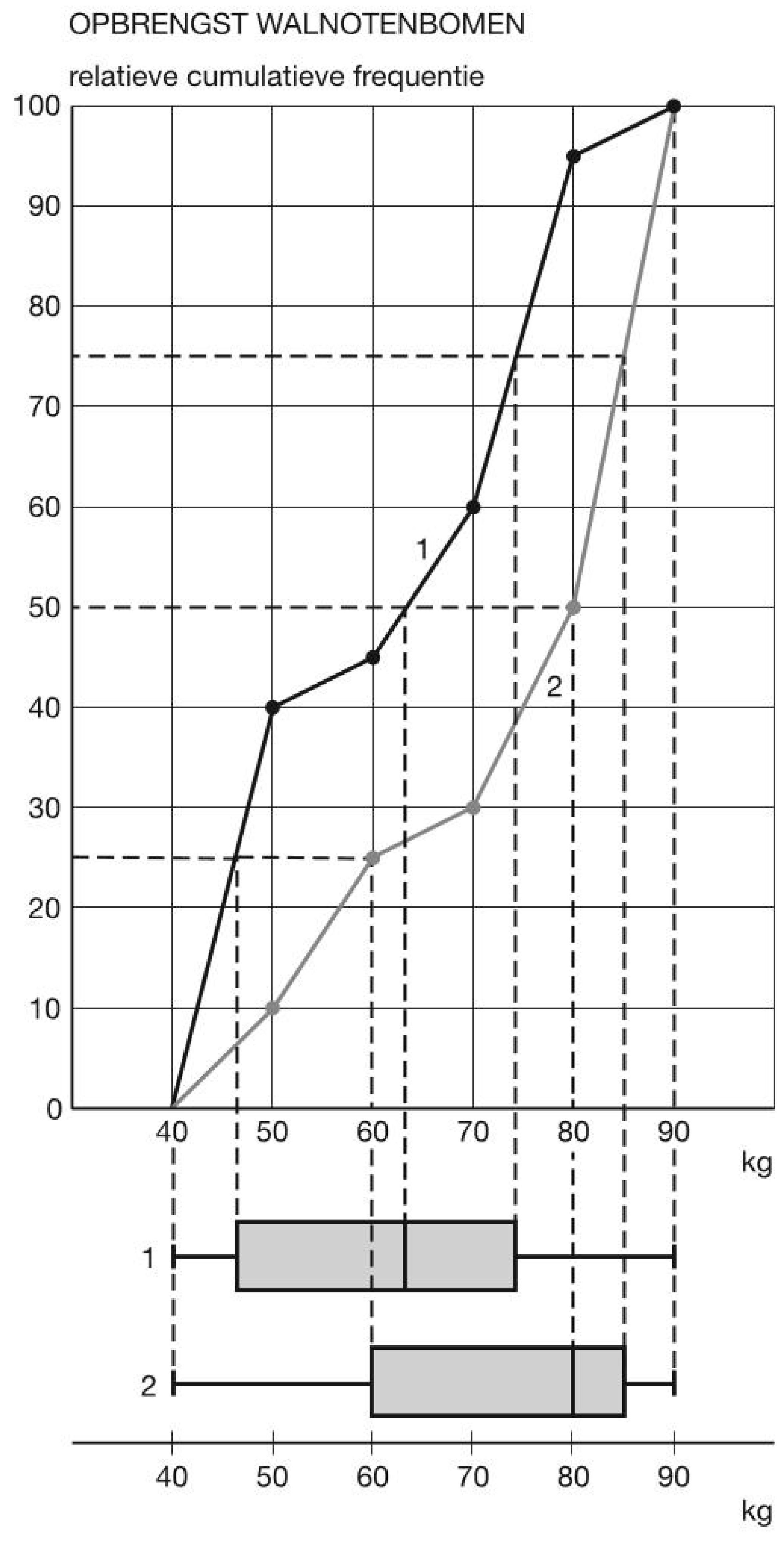
c • De boxen overlappen elkaar niet.
Dus het verschil is groot.

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

14 **a** Zie de figuur hiernaast.
• De boxen overlappen elkaar.
• Er is een mediaan die buiten de box van de andere boxplot ligt.
Dus het verschil is middelmatig.

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

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



10

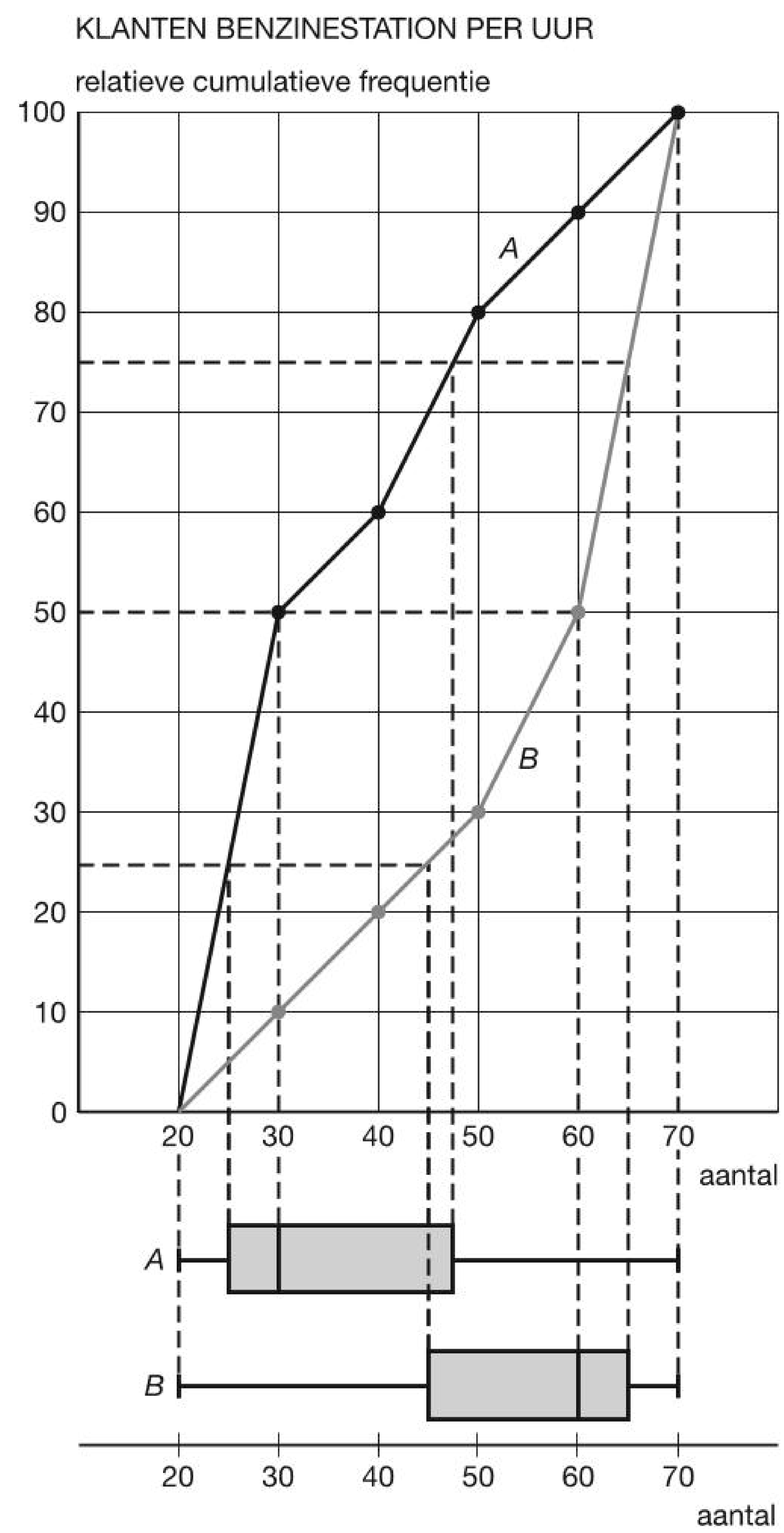
Bladzijde 70

15 **a** Bij *B* was het het drukst. De grafiek van *B* ligt geheel rechts van de grafiek van *A*, dus bij *B* horen bij hogere aantallen klanten per uur grotere percentages.

b De *Vcp* is maximaal bij 50 klanten per uur.
$max.Vcp = 80\% - 30\% = 50\%$.
Omdat $max.Vcp > 40\%$ is het verschil groot.

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

Dus het verschil is middelmatig.
Je trekt niet twee keer dezelfde conclusie.



16 **a**

	Arnhem		Nijmegen			
	cum.	cum. perc.	cum.	cum. perc.	*Vcp*	
zeer mee oneens	5	4,0%	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%	← *max.Vcp*
zeer mee eens	125	100%	120	100%	0%	

$max.Vcp = 24{,}5\%$
Omdat $20\% < max.Vcp \leq 40\%$ is het verschil middelmatig.

b

	Arnhem cum. perc.	Nijmegen cum. perc.	*Vcp*	
(zeer) mee oneens	8,8%	3,3%	5,5%	← *max.Vcp*
neutraal	18,4%	18,3%	0,1%	
(zeer) mee eens	100%	100%	0%	

$max.Vcp = 5{,}5\%$
Omdat $max.Vcp \leq 20\%$ is het verschil gering.
Je trekt dus niet dezelfde conclusie.

Bladzijde 71

17 **a** I is juist. Als de teller toeneemt en de noemer blijft gelijk dan neemt de breuk toe.
II is onjuist. Als de noemer toeneemt en de teller blijft gelijk dan neemt de breuk juist af.

III is juist, want $\dfrac{1{,}10\overline{X}_1 - 1{,}10\overline{X}_2}{\frac{1}{2}(S_1 + S_2)} = \dfrac{1{,}10(\overline{X}_1 - \overline{X}_2)}{\frac{1}{2}(S_1 + S_2)} = 1{,}10 \cdot \dfrac{\overline{X}_1 - \overline{X}_2}{\frac{1}{2}(S_1 + S_2)} = 1{,}10E$.

IV is juist, want $\dfrac{\overline{X}_1 - \overline{X}_2}{\frac{1}{2}(2S_1 + 2S_2)} = \dfrac{(\overline{X}_1 - \overline{X}_2)}{\frac{1}{2} \cdot 2(S_1 + S_2)} = \dfrac{1}{2} \cdot \dfrac{\overline{X}_1 - \overline{X}_2}{\frac{1}{2}(S_1 + S_2)} = \frac{1}{2}E$.

b $E = \dfrac{3673 - 3497}{\frac{1}{2}(442 + 428)} \approx 0{,}41$

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

c • De boxen overlappen elkaar.
• 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.

10.3 De statistische cyclus

Bladzijde 73

18 **a** Bijvoorbeeld ‘aantal vorstdagen’, met de deelvraag ‘Is het aantal vorstdagen vanaf 1980 kleiner dan voor 1980?’.

b Ja, alle genoemde jaren zijn vanaf 1980. Van voor 1980 staan er geen jaren in de top-10.
Bij de grootste neerslagsom zijn 5 van de jaren uit de top-10 van voor 1980 en 5 van de jaren vanaf 1980. Dit geeft geen aanleiding om een bevestigend antwoord te geven op deelvraag 2.
Bij het grootste aantal uren zonneschijn zijn er 4 jaren van voor 1980 en 6 van erna. Dat geeft onvoldoende aanleiding een bevestigend antwoord te geven op deelvraag 3.

Bladzijde 75

19 $E = \dfrac{11{,}07 - 10{,}10}{\frac{1}{2}(0{,}92 + 0{,}73)} \approx 1{,}18$

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

20 **a** De vraag is suggestief. Dit is dus geen goede vraag.

b Nee, de steekproef is niet aselect. Alleen leerlingen en docenten van deze school zijn geënquêteerd.

c $phi = \dfrac{18 \cdot 25 - 7 \cdot 23}{\sqrt{25 \cdot 48 \cdot 41 \cdot 32}} \approx 0{,}23$

Omdat $0{,}2 < phi < 0{,}4$ is het verschil middelmatig.

21 $E = \dfrac{558{,}0 - 523{,}0}{\frac{1}{2}(115{,}5 + 99{,}6)} \approx 0{,}33$

Omdat $E < 0{,}4$ is het verschil gering.

Bladzijde 76

22 **a** Stel: in een zomer is op 10 dagen de gemiddelde etmaaltemperatuur 20 °C en op 10 dagen 12 °C.
De 10 dagen met een temperatuur van 20 °C dragen $10 \cdot (20 - 18) = 10 \cdot 2 = 20$ bij aan het warmtegetal terwijl de 10 dagen met een temperatuur van 12 °C niets bijdragen aan maar ook niets afnemen van het warmtegetal.

b Het warmtegetal zegt alleen iets over de periode tussen 1 april en 31 oktober, terwijl de gemiddelde temperaturen over een jaar zijn berekend.

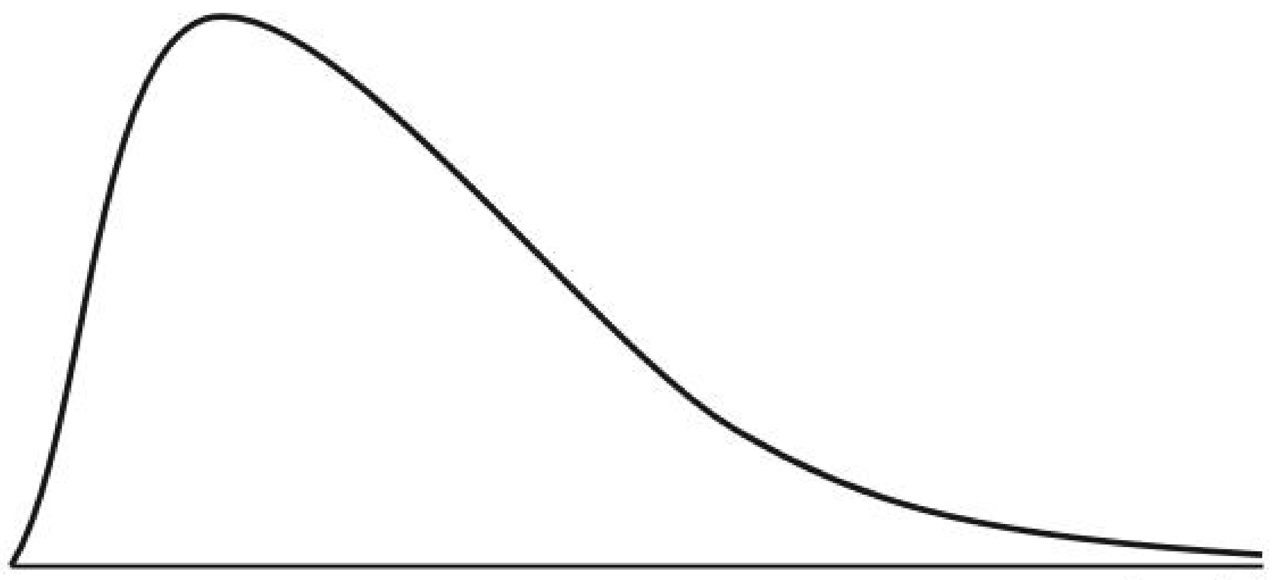
c

dag	3	5	6	7	8	9	10	11	12
gem. temperatuur in °C	21,5	18,9	20,5	20,6	18,7	18,6	19,8	20	19,5
bijdrage aan warmtegetal	3,5	0,9	2,5	2,6	0,7	0,6	1,8	2	1,5

dag	13	14	15	21	22	23	26	30	31
gem. temperatuur in °C	21,7	20,9	18,4	19,7	20,7	19,7	20,2	20	20,9
bijdrage aan warmtegetal	3,7	2,9	0,4	1,7	2,7	1,7	2,2	2	2,9

Het warmtegetal is $3{,}5 + 0{,}9 + 2{,}5 + 2{,}6 + 0{,}7 + 0{,}6 + 1{,}8 + 2 + 1{,}5 + 3{,}7 + 2{,}9 + 0{,}4 + 1{,}7 + 2{,}7 + 1{,}7 + 2{,}2 + 2 + 2{,}9 = 36{,}3$.
Het warmtegetal dat bij augustus 2015 hoort is dus ongeveer 36.

d De mediaan is kleiner dan het gemiddelde, dus een rechts-scheve verdeling.



e 12 van de 14 jaren van de periode 2001-2014 hebben een warmtegetal dat groter is dan het gemiddelde. En ook 12 van de 14 jaren hebben een warmtegetal dat groter is dan de mediaan. Op zichzelf is dit te weinig om een conclusie uit te trekken over klimaatverandering, maar het wijst wel op het warmer worden van de aarde.

f $E = \dfrac{94{,}2 - 58{,}5}{\frac{1}{2}(36{,}4 + 37{,}4)} \approx 0{,}967$

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

Bladzijde 78

23 $E = \dfrac{1311 - 1256}{\frac{1}{2}(136 + 132)} \approx 0{,}410$

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

24 **a** 1 Kan kloppen.
De spreiding van de 50% jaren met de kleinste aantallen zonuren is kleiner dan de 50% jaren met de grootste aantallen zonuren.
2 Kan niet kloppen.
Van de jaren 1901-1979 heeft 25% meer dan 1340 uren zon per jaar en $0{,}25 \cdot 79$ is minder dan 40 jaren.
3 Kan kloppen.
In de boxplot van de periode 1980-2014 lees je af dat aantallen zonuren tussen 1300 en 1400 bij 25% tot 50% van deze jaren voorkwamen. Dus bij minstens $0{,}25 \cdot 35 = 8{,}75$ jaren en hoogstens $0{,}5 \cdot 35 = 17{,}5$ jaren.
4 Kan niet kloppen.
In de periode 1901-1979 is het aantal jaren met tussen 1250 en 1400 uren zon meer dan 25% van 79.
In de periode 1980-2014 is dit aantal jaren 50% van 35.
Omdat 25% van 79 al groter is dan 50% van 35 kan dit niet juist zijn.

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

Dus het verschil is gering.

25 **a** In deze opgaven zijn steeds de jaren voor 1980 vergeleken met de jaren vanaf 1980.
De resultaten:
- (opgave 19) Het verschil in temperatuur is groot (vanaf 1980 gemiddeld hogere temperatuur).
- (opgave 21) Het verschil in hoeveelheidneerslag is gering.
- (opgave 23) Het verschil in aantal uren zon is middelmatig (vanaf 1980 gemiddeld meer zon).

Op grond van deze resultaten kun je concluderen dat het klimaat in Nederland na 1900 veranderd is. Dit komt vooral tot uiting in de temperatuur maar ook in het aantal uren zon is de klimaatverandering zichtbaar. De verandering van de hoeveelheid neerslag is gering.

b *

10.4 Interpreteren van onderzoeksresultaten

Bladzijde 79

26 **a** De procentuele verandering is $\dfrac{154 - 164}{164} \times 100\% \approx -6{,}1\%$,
dus het aantal miljonairshuishoudens is met 6,1% afgenomen.

b Nee, daarvoor heb je het gemiddelde in plaats van het mediane vermogen nodig.

10

c De 10% laagste inkomens hadden 7% van het totale vermogen, dat was $0{,}07 \cdot 1166$ miljard = 81,62 miljard euro. Deze $0{,}1 \cdot 7410$ duizend = 741 duizend huishoudens hadden een gemiddeld vermogen van $\frac{81{,}62 \cdot 10^9}{741 \cdot 10^3} \approx 110\,148$ euro per huishouden.

d Met ‘doorsnee’ wordt niet het gemiddelde bedoeld.

Bladzijde 80

27 **a** $P_v = 2$ en $P_k = 4$ geeft $E = (2 + 0{,}8 \cdot 4)^{0{,}5} \approx 2{,}28$.

b $P_v = 2$ en $P_k = 2$ geeft $E = (2 + 0{,}8 \cdot 2)^{0{,}5} = 1{,}897...$ voor huishouden I.
$P_v = 3$ en $P_k = 3$ geeft $E = (3 + 0{,}8 \cdot 3)^{0{,}5} = 2{,}323...$ voor huishouden II.
Huishouden II heeft, om even goed rond te komen, $\frac{2{,}323...}{1{,}897...} = 1{,}224...$ keer zo veel besteedbaar inkomen nodig als huishouden I, dat is dus € $45\,000 \cdot 1{,}224... \approx 55\,000$ euro.

28 I Deze conclusie kun je trekken.
In de tabel staat dat 96,1% van de huishoudens een primair inkomen heeft.

II Deze conclusie kun je trekken.
In het diagram zie je dat bij de 6^e tot en met de 10^e 10%-groep het inkomen steeds uit meer dan 50% primair inkomen bestaat.

III Deze conclusie kun je niet trekken.
In het diagram is te zien dat van het inkomen van de 10% huishoudens met het hoogste bruto inkomen, ongeveer 94% uit primair inkomen bestaat. Of dit primair inkomen voor deze groep bestaat uit (onder andere) inkomen uit eigen onderneming is niet uit deze gegevens te halen. Het kan ook uitsluitend bestaan uit inkomen uit arbeid en/of inkomen uit vermogen.

IV Deze conclusie kun je niet trekken.
In het diagram is te zien dat hoe hoger het inkomen is, hoe minder van het inkomen bestaat uit uitkeringen. Het inkomen van de 1^e, 2^e, 3^e en 4^e 10%-groep bestaat voor meer dan de helft uit uitkeringen, maar bij de 5^e 10%-groep bestaat het inkomen voor minder dan 50% uit uitkeringen.

V Deze conclusie kun je trekken.
In de kolom ‘gemiddeld bedrag’ van de tabel staat dat het bruto-inkomen gemiddeld 57,4 duizend euro is (dit bedrag is ook af te lezen in het diagram bij totaal).
Uit het diagram valt op te maken dat de mediaan tussen 40,8 duizend en 50,8 duizend ligt.

Bladzijde 82

29 **a** Je hebt te maken met een rechts-scheve verdeling.

b De modale klasse is 16 000 –< 18 000 euro. De mediaan ligt hier rechts van.
De mediaan ligt links van het gemiddelde want de verdeling is rechts-scheef.

c II, veel lage inkomens en weinig hoge dus de cumulatieve verdelingskromme stijgt eerst sterk en daarna nog maar weinig.

10

Bladzijde 83

30 **a** In elke groep zitten 741 000 huishoudens.

$$\text{Totale vermogen} = (113 + 60 + 73 + 94 + 117 + 130 + 142 + 169 + 217 + 459) \cdot 1000 \cdot 741\,000$$
$$\approx 1{,}166 \cdot 10^{12} = 1166 \cdot 10^9 = 1166 \text{ miljard euro.}$$

b I Deze uitspraak is te verdedigen.
De hoogste 10%-groep bezit $\frac{741\,000 \cdot 459\,000}{1166 \cdot 10^9} \times 100\% \approx 29{,}2\%$, dus bijna 30% van het vermogen.

II Deze uitspraak is niet te verdedigen.
De mediaan van deze groep is 30 000 euro. Dus de helft van 741 000 huishoudens uit de 5^e 10%-groep heeft een vermogen van meer dan 30 000 euro, dat zijn er $0{,}5 \cdot 741\,000 = 370\,500$.
Dat is dus minder dan 400 duizend huishoudens.

III Deze uitspraak is niet te verdedigen.
Bij een links-scheve verdeling is de mediaan groter dan het gemiddelde maar hier is de mediaan kleiner dan het gemiddelde.

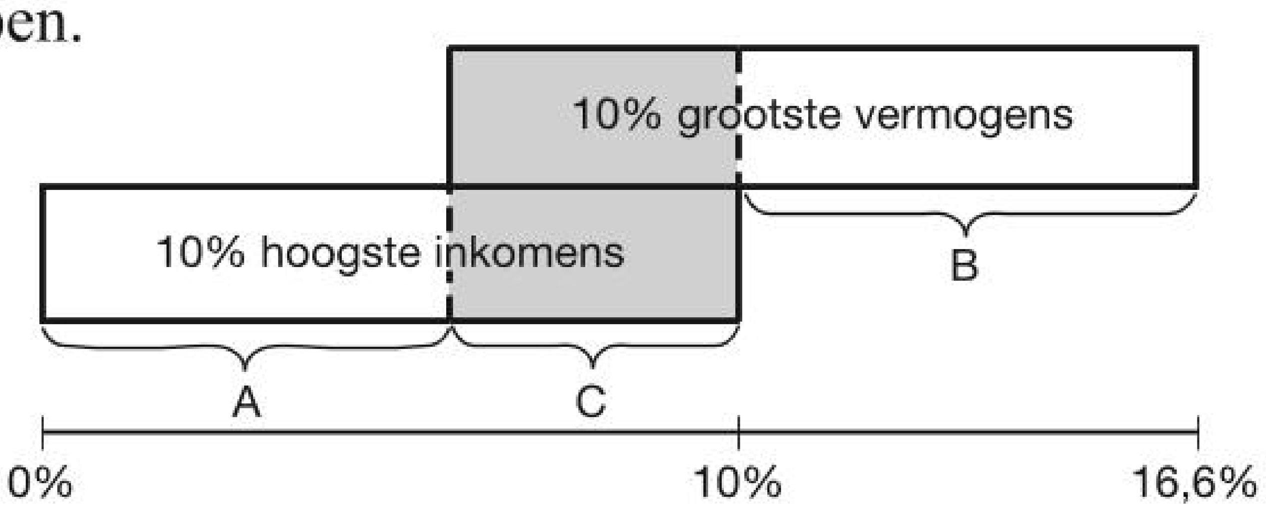
IV Deze uitspraak is te verdedigen.
In de tabel is te zien dat ruwweg geldt: hoe hoger het (gemiddelde/mediane) inkomen hoe groter de omvang van hethuishouden.

31 **a** Als de 10% huishoudens met de hoogste gestandaardiseerde inkomens precies dezelfde huishoudens zijn als de 10% huishoudens met de grootste vermogens dan is het aantal huishoudens met een hoge welvaart 10%.
Als geen van de huishoudens die behoort tot de 10% huishoudens met de hoogste gestandaardiseerde inkomens ook behoort tot de 10% huishoudens met de grootste vermogens (de groepen overlappen elkaar niet) dan is het percentage huishoudens met een hoge welvaart 20%.
Als er zowel huishoudens zijn die tot beide groepen behoren als huishoudens die tot slechts één van beide 10%-groepen behoren (de groepen overlappen gedeeltelijk) dan ligt het percentage huishoudens met hoge welvaart dus tussen 10% en 20%.

b Stel:

- A is het percentage huishoudens dat behoort tot de 10% huishoudens met de hoogste gestandaardiseerde inkomens maar niet behoort tot de 10% huishoudens met de grootste vermogens.
- B is het percentage huishoudens dat behoort tot de 10% huishoudens met de grootste vermogens maar niet tot de 10% huishoudens met de hoogste gestandaardiseerde inkomens.
- C is het percentage huishoudens dat behoort tot beide 10%-groepen.

Dan geldt $A + B + C = 16{,}6\%$.
Er moet ook gelden $B + C = 10\%$ en ook $A + C = 10\%$.



$$\left.\begin{array}{l} A + B + C = 16{,}6\% \\ B + C = 10 \end{array}\right\} \begin{array}{l} A + 10\% = 16{,}6\% \\ A = 6{,}6\% \end{array}$$

$A = 6{,}6\%$ en $A + C = 10\%$ geeft $6{,}6\% + C = 10\%$ ofwel $C = 3{,}4\%$.
Dus 3,4% behoort zowel tot de 10% huishoudens met de hoogste gestandaardiseerde inkomens als tot de 10 procent huishoudens met de grootste vermogens.

Bladzijde 84

32 **a** Elke minuut is een klasse, dus 60 klassen.

b Nee, alle klassen bevatten evenveel huishoudens. Langs de verticale as staat niet het aantal huishoudens.

c Bij huishoudens met een inkomen van minstens 50 duizend euro horen de laatste 2 minuten.
Dat is dus $\frac{2}{60} \times 100\% \approx 3{,}3\%$.

d De hoogte van de staaf bij de zestigste minuut wordt niet bepaald door het hoogste inkomen, maar is een gemiddelde van de inkomens in die klasse.

e De relatieve cumulatieve frequentie van 500 000 euro is 93%. Dus $100\% - 93\% = 7\%$ heeft een vermogen van minstens een half miljoen euro.

f Bij een half uur hoort de relatieve cumulatieve frequentie 50% en hierbij hoort een vermogen van ongeveer 27 500 euro.

g Bij 157,5 duizend euro is de relatieve cumulatieve frequentie ongeveer 73%.
$0{,}73 \cdot 60 = 43{,}8$, dus in de 44e minuut.

h Bij het gemiddelde van 157,5 duizend euro hoort de lengte 1,74 meter.
Eén miljoen is $\frac{1 \cdot 10^6}{157{,}5 \cdot 10^3} = 6{,}349...$ keer zo groot als het gemiddelde vermogen.
Dus bij één miljoen hoort de lengte $6{,}349... \cdot 1{,}74 \approx 11{,}05$ m.

i Bij het gemiddelde van 157,5 duizend euro hoort de lengte 1,74 meter.
30,5 meter is $\frac{30{,}5}{1{,}74} = 17{,}528...$ keer zo groot als 1,74 meter.
Dus bij 30,5 meter hoort een vermogen van $17{,}528... \cdot 157{,}5 \approx 2761$ duizend euro $\approx 2{,}76$ miljoen euro.

Bladzijde 86

33 I Deze uitspraak is te verdedigen.
Je ziet in de tabel dat geldt: hoe hoger de inkomensgroep, hoe kleiner het percentage van het besteedbaar inkomen wordt uitgegeven aan huisvesting, water en energie.

II Deze uitspraak is niet te verdedigen.
Het percentage bij de alleenstaande is hoger dan bij het eenoudergezin, maar dit zegt niets over het bedrag omdat deze percentages niet van hetzelfde bedrag zijn.

III Deze uitspraak is te verdedigen.
Naar recreatie en cultuur gaat 12,0% van het geld en naar vervoer gaat 7,1% van het geld.
Er wordt $\frac{12{,}0 - 7{,}1}{7{,}1} \times 100\% \approx 69\%$, dus bijna 70% meer aan recreatie en cultuur uitgegeven dan aan vervoer.

10

IV Deze uitspraak is niet te verdedigen.

Hoewel $\frac{18,2}{12,0} \approx 1,52$ zegt deze verhouding niets over de verhouding van de bedragen die besteed worden aan recreatie en cultuur. Het totale bedrag dat deze groepen te besteden hebben is namelijk niet gelijk. Het percentage 18,2 is van een hoger bedrag dan het percentage 12,0.

Bladzijde 87

34 **a** In 2006 was het totale bedrag dat aan zorgtoeslag werd uitgekeerd
$4,0 \cdot 10^6 \cdot 550 = 2200 \cdot 10^6 = 2,2 \cdot 10^9 = 2,2$ miljard euro.
In 2012 was dat $4,6 \cdot 10^6 \cdot 980 = 4508 \cdot 10^6 \approx 4,5 \cdot 10^9 = 4,5$ miljard euro.
De procentuele toename is $\frac{4,5 - 2,2}{2,2} \times 100\% \approx 104,5\%$.

b De percentages huishoudens met zorgtoeslag bij de eerste drie 10%-groepen zijn 83%, 92% en 86%.
Omdat de groepen even groot zijn ontving $\frac{83\% + 92\% + 86\%}{3} = 87\%$ van deze huishoudens zorgtoeslag.

c In figuur 10.16 is af te lezen dat 62% van alle huishoudens zorgtoeslag ontving, dat waren volgens figuur 10.15 4,6 miljoen huishoudens. Dus het totaal aantal huishoudens is $\frac{4,6}{0,62} \approx 7,4$ miljoen.
In 2012 werd 4508 miljoen euro zorgtoeslag uitgekeerd (zie vraag a).
Het gemiddelde bedrag per huishouden was dus $\frac{4508}{7,4} \approx 609$ euro.

Diagnostische toets

Bladzijde 88

1 **a** $\hat{p} = \frac{80}{500} = 0,16$ en $\sigma = \sqrt{\frac{0,16 \cdot 0,84}{500}} = 0,0163...$

$\hat{p} - 2\sigma = 0,16 - 2 \cdot 0,0163... \approx 0,127$
$\hat{p} + 2\sigma = 0,16 + 2 \cdot 0,0163... \approx 0,193$
Het 95%-betrouwbaarheidsinterval is [0,127; 0,193].

b $4\sigma = 0,176 - 0,099 = 0,077$, dus $\sigma = \frac{0,077}{4} = 0,01925$

$\hat{p} = \frac{0,099 + 0,176}{2} = 0,1375$, dus $\sigma = \sqrt{\frac{0,1375 \cdot 0,8625}{n}}$

Los op $\sqrt{\frac{0,1375 \cdot 0,8625}{n}} = 0,01925$.

Voer in $y_1 = \sqrt{\frac{0,1375 \cdot 0,8625}{n}}$ en $y_2 = 0,01925$.
Intersect geeft $x \approx 320$.
Dus aan 320 studenten is gevraagd of zij rood stonden.
Hiervan stonden er $0,1375 \cdot 320 \approx 44$ rood en de helft hiervan, dus 22 studenten, stonden meer dan 700 euro rood.

2 $\hat{p} = \frac{195}{1500} = 0,13$ en $\sigma = \sqrt{\frac{0,13 \cdot 0,87}{1500}} = 0,0086...$

$\hat{p} - 2\sigma = 0,13 - 2 \cdot 0,0086... \approx 0,1126...$ en $0,1126... \cdot 250\,000 \approx 28\,158$
en $0,1473... \cdot 250\,000 \approx 36\,062$
Het 95%-betrouwbaarheidsinterval voor het aantal buitenlandse studenten is [28 158; 36 062].

10

3 **a** Het 95%-betrouwbaarheidsinterval is $\left[48{,}4 - 2 \cdot \frac{11{,}2}{\sqrt{300}}, 48{,}4 + 2 \cdot \frac{11{,}2}{\sqrt{300}}\right] \approx [47{,}11; 49{,}69]$.

b $4 \cdot \frac{S}{\sqrt{300}} = 17{,}73 - 16{,}27 = 1{,}46$

Voer in $y_1 = 4 \cdot \frac{x}{\sqrt{300}}$ en $y_2 = 1{,}46$.
Intersect geeft $x \approx 6{,}3$.
Dus de steekproefstandaardafwijking is 6,3 kg.

4 $PV = \frac{66}{126} \times 100\% - \frac{34}{91} \times 100\% \approx 15{,}02\%$
Omdat $15\% < PV < 30\%$ is het verschil middelmatig.
$OR = \frac{57 \cdot 66}{34 \cdot 60} \approx 1{,}8$
Omdat $OR < 2$ is het verschil gering.
$phi = \frac{34 \cdot 60 - 57 \cdot 66}{\sqrt{91 \cdot 126 \cdot 100 \cdot 117}} \approx -0{,}149$
Omdat $-0{,}2 < phi < 0{,}2$ is het verschil gering.
Omdat het *PV* maar net boven 15% zit en zowel de *OR* als *phi* een gering verschil aangeven, trekken we de conclusie dat het verschil gering is.

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

Omdat $E > 0{,}8$ is dit verschil groot.

Bladzijde 89

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

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

Verschil tussen A en C:
De *Vcp* is maximaal bij 120 seconden.
$max.Vcp = 90\% - 45\% = 45\%$.
Omdat $max.Vcp > 40\%$ 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.

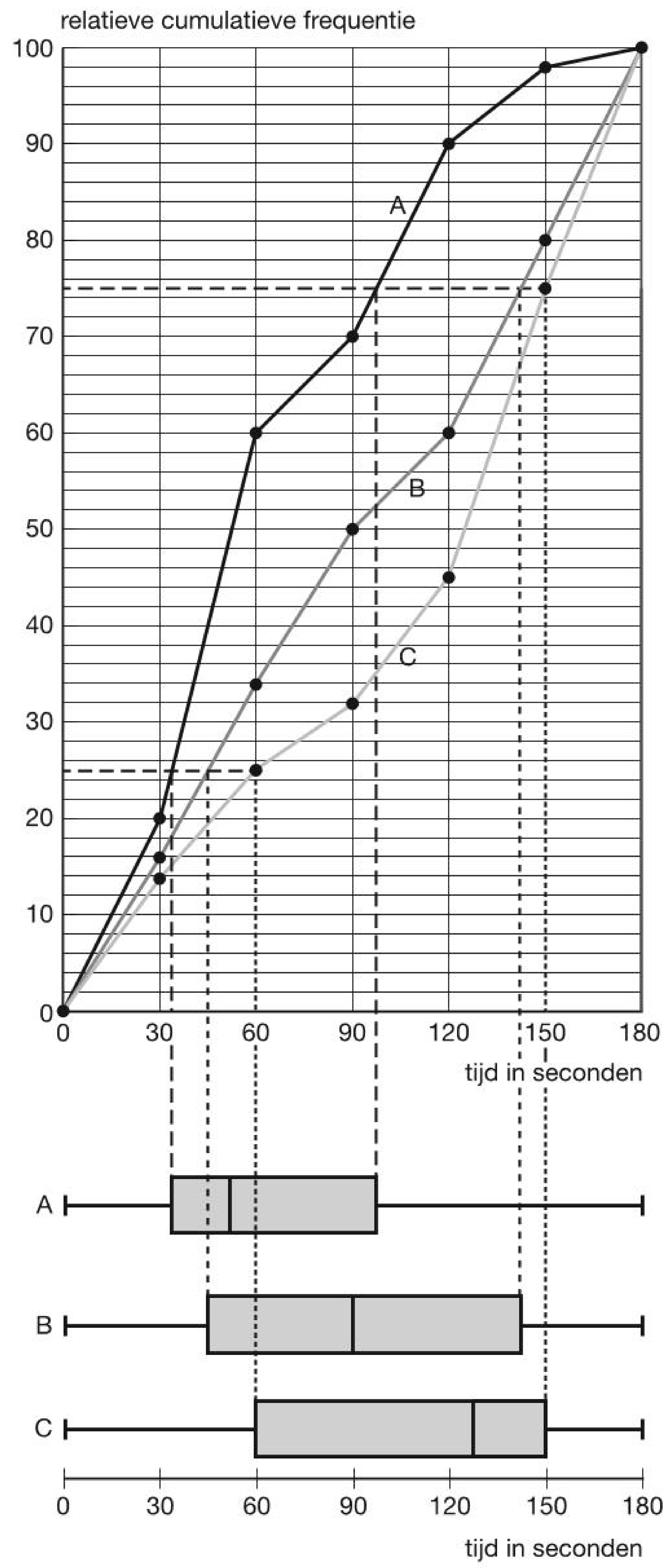
Dus het verschil is gering.
Dit is niet dezelfde conclusie als bij vraag b.

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.



10

7 **a** Een mogelijke hoofdvraag is:
In welke mate is er een verschil in lichaamsbouw tussen wezels uit het noorden van Europa en uit het zuiden van Europa?
Bijbehorende deelvragen kunnen zijn:
In welke mate is er een verschil in gewicht tussen wezels uit het noorden van Europa en uit het zuiden van Europa?
In welke mate is er een verschil in lichaamslengte tussen wezels uit het noorden van Europa en uit het zuiden van Europa?

b $E = \dfrac{54 - 49}{\frac{1}{2}(5{,}7 + 5{,}3)} \approx 0{,}909$

Omdat $E > 0{,}8$ is dit verschil groot.

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

Dus het verschil is middelmatig.

d $$phi = \dfrac{438 \cdot 336 - 242 \cdot 294}{\sqrt{(438 + 242) \cdot (294 + 336) \cdot (438 + 294) \cdot (242 + 336)}}$$
$$= \dfrac{76020}{\sqrt{680 \cdot 630 \cdot 732 \cdot 578}} \approx 0{,}179$$

Omdat $-0{,}2 < phi < 0{,}2$ is het verschil gering.

e Het verschil in gewicht is groot en het verschil in kop-romplengte is middelmatig maar komt wel in de buurt van groot. Het verschil in staartlengte is weliswaar gering maar is een minder belangrijke factor in de forsheid van een dier. Al met al is het verschil middelmatig te noemen.

10