

10 Statistiek gebruiken

Voorkennis Statistische variabelen

Bladzijde 54

- 1** geboortejaar: Wel getallen, wel hoeveelheden, niet alle tussenliggende waarden mogelijk.
Dus discreet.
hoogst genoten opleiding: Geen getallen, wel duidelijke ordening.
Dus ordinaal.
telefoonnummer: Wel getallen, geen hoeveelheden, geen duidelijke ordening.
Dus nominaal.
mogen we u mail sturen?: Geen getallen, geen duidelijke ordening.
Dus nominaal.
- 2** **a** Geen getallen, geen duidelijke ordening. Dus nominaal.
b Wel getallen, wel hoeveelheden, alle tussenliggende waarden mogelijk. Dus continu.
c Wel getallen, wel hoeveelheden, niet alle tussenliggende waarden mogelijk. Dus discreet.
d Geen getallen, wel duidelijke ordening. Dus ordinaal.
e Geen getallen, geen duidelijke ordening. Dus nominaal.
f Wel getallen, wel hoeveelheden, niet alle tussenliggende waarden mogelijk. Dus discreet.
- 3** Gele kaarten: Wel getallen, wel hoeveelheden, niet alle tussenliggende waarden mogelijk.
Dus kwantitatief en discreet.
Zorgen nalatenschap: Geen getallen, wel duidelijke ordening.
Dus kwalitatief en ordinaal.
BMI: Wel getallen, wel hoeveelheden, alle tussenliggende waarden mogelijk.
Dus kwantitatief en continu.
Verdeling bloedgroepen: Geen getallen, geen duidelijke ordening.
Dus kwalitatief en nominaal.

10.1 Statistische diagrammen

Bladzijde 55

- 1** **a** Nee, dat kan niet, omdat het verschil tussen de waarden 'man' en 'vrouw' geen afstand is. Je kunt ze dus niet uitzetten op de horizontale as.
b Ja, dat kan. Je krijgt een cirkeldiagram met twee sectoren, namelijk man en vrouw.

Bladzijde 56

- 2** **a** ordinaal
b staafdiagram en cirkeldiagram
c Vanwege de klassen $\langle \leftarrow; 1,70 \rangle$ en $\langle 1,90; \rightarrow \rangle$. Die hebben geen klassenmidden.

Bladzijde 57

- 3** **a** ordinaal
b Bij de taalvakken: $100 - 6,8 = 93,2\%$.
Bij de exacte vakken: $100 - 10,7 = 89,3\%$.
c Van de centrummaten alleen de modus.
Geen enkele spreidingsmaat.
d Een cirkeldiagram. De data is kwalitatief, dus naast een stapeldiagram (of staafdiagram) is alleen een cirkeldiagram mogelijk.

- 4** a De variabele geslacht. Het meetniveau is nominaal.
 b Hij koos voor groep 1 de jongens.
 Bij de jongens komen grotere aantallen (boven 20) relatief vaker voor dan bij de meisjes, en lagere aantallen (onder 20) juist relatief minder vaak. Het gemiddelde is dus bij de jongens groter dan bij de meisjes.
 Bij de meisjes is de spreiding van de waarnemingsgetallen groter dan bij de jongens, dus bij de meisjes is de standaardafwijking groter dan bij de jongens.
 c Centrummaat: de mediaan. (Niet de modus, want die is er niet bij de jongens.)
 Spreidingsmaten: spreidingsbreedte en interkwartielafstand.
 d Nee. Voor een puntenwolk moet per leerling het aantal dagelijks beluisterde liedjes en de leeftijd bekend zijn.

Bladzijde 58

- 5** a De variabele dagtemperatuur is kwantitatief.
 Het meetniveau van de variabele maand is ordinaal.
 b Je ziet dat de blauwe boxplots doorgaans lager liggen dan de rode boxplots. De rode boxplots horen dus bij periode II.
 c interkwartielafstand = $7 - 2,5 = 4,5$
 Dus er is sprake van een uitschieter bij een temperatuur van minder dan $2,5 - 1,5 \cdot 4,5 = -4,25$ °C en meer dan $7 + 1,5 \cdot 4,5 = 13,75$ °C.
 d Bijvoorbeeld januari in periode II, want deze boxplot hoort bij een links-scheve verdeling.
 e Bij periode I wordt de spreidingsbreedte $22,5 - -9 = 31,5$ °C en bij periode II wordt de spreidingsbreedte $26 - -4 = 30$ °C. Dus de grootste spreidingsbreedte is 31,5 °C.

Bladzijde 59

- 6** a Uit de cumulatieve frequentiepolygoon lees je af dat er 350 dagen in zijn verwerkt.
 b De rode lijndiagrammen horen bij de kamerverhuur van 2019, want bij deze lijndiagrammen komen grotere aantallen verhuurde kamers minder voor.
 c In 2015 is de mediaan ongeveer 51 en in 2019 is de mediaan ongeveer 38.
 In 2015 is de modale klasse $[60, 70)$ en in 2019 is de modale klasse $[40, 50)$.
 d De mediaan 40 (bij de rode cumulatieve frequentiepolygoon) kan niet kloppen, want de cumulatieve frequentie 175 valt in de klasse $[30, 40)$.
 De mediaan 53 (bij de blauwe cumulatieve frequentiepolygoon) kan wel kloppen, want de cumulatieve frequentie 175 valt in de klasse $[50, 60)$.
 e Bert heeft gelijk. Je leest bij de relatieve polygoon bij het klassenmidden 55 af dat de relatieve frequenties niet gelijk zijn. Bij de rode polygoon is de relatieve frequentie minder dan 15 en bij de blauwe polygoon is dat meer dan 15. En omdat het aantal waarnemingsgetallen voor beide gelijk is, namelijk 350, zijn de frequenties niet gelijk.

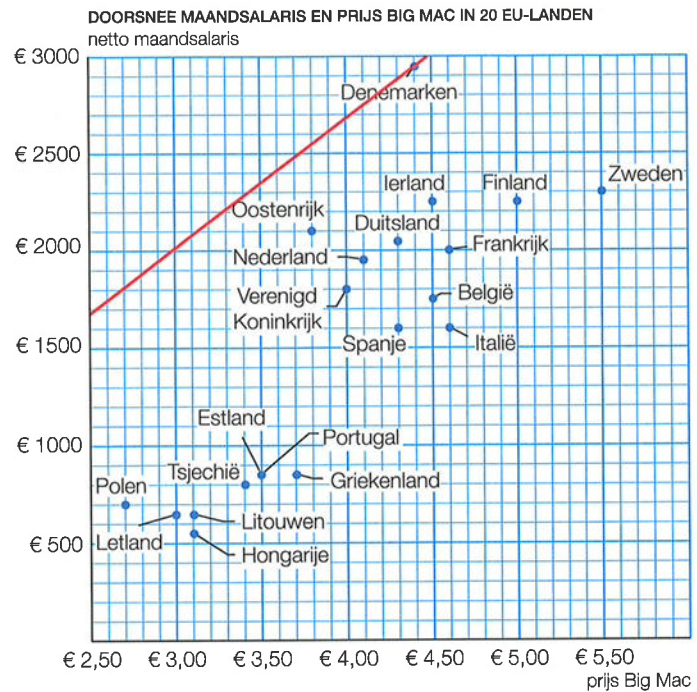
Bladzijde 60

- 7** a spreidingsdiagram
 b Estland en Portugal horen bij hetzelfde punt.
 c Het gemiddelde is niet geschikt. Omdat de verdeling van de maandsalarissen rechts-scheef is, geeft het gemiddelde maandsalaris een te hoge waarde om nog van een doorsnee maandsalaris te kunnen spreken.

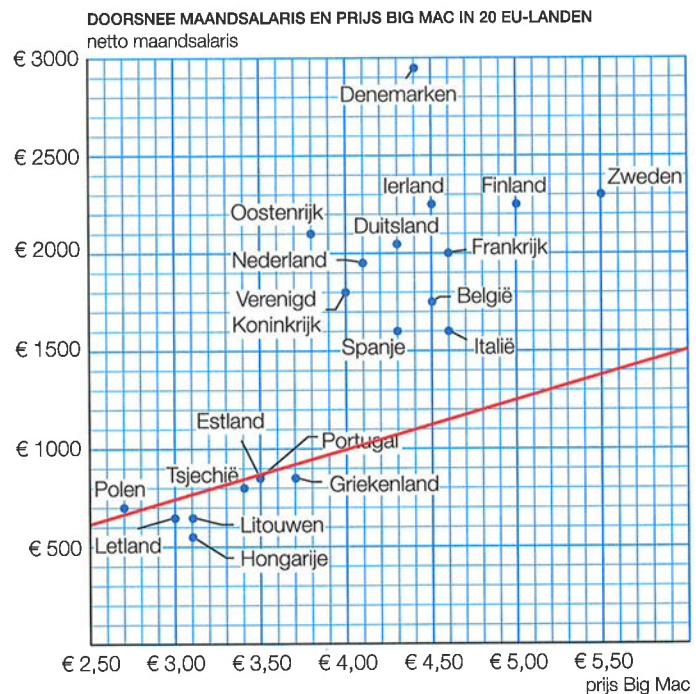
d kwantitatief

e Omdat Denemarken een hoog netto maandsalaris heeft en de prijs van een Big Mac daar niet al te hoog is, komt Denemarken in aanmerking. In Denemarken kun je voor een netto maandsalaris ongeveer $\frac{2950}{4,40} \approx 670$ Big

Macs kopen. Bij een prijs van 3 euro zou je dus een netto maandsalaris van $3 \cdot 670 = 2010$ euro moeten hebben om evenveel Big Macs te kunnen kopen. Teken daarom de lijn door het punt (3, 2010) en het punt van Denemarken. Er liggen geen punten boven deze lijn, dus het antwoord is Denemarken.



f Teken de lijn door de punten (3, 750) en (5, 1250). De punten van de landen onder deze lijn zijn de landen waar je minder dan 250 keer een Big Mac van het netto maandsalaris kunt kopen. Dit zijn de landen Letland, Hongarije, Litouwen, Tsjechië, Estland, Portugal en Griekenland.



Bladzijde 61

- 8 a dotplot
b Bij totaal minstens 10 optie 2. Bij niet minder dan in totaal 4 optie 1.
c Bij de D6-worpen hoort een normale verdeling. Bij de D12-worp hoort een uniforme verdeling.
d Bij de D6-worpen. Daar zijn veel uitkomsten rond het gemiddelde en relatief weinig lage en hoge uitkomsten.
- 9 Uitgaande van 100 sporters wonnen $0,50 \cdot 70 + 0,60 \cdot 30 = 53$ een medaille. Hierbij waren $0,50 \cdot 70 = 35$ vrouwen.
Dus $\frac{35}{53} \times 100\% \approx 66\%$ van de medaillewinnaars is vrouw.

Bladzijde 62

- 10 a** ordinaal
b staafdiagram of cirkeldiagram
c Je kunt alleen bij hoeveelheden een gemiddelde berekenen en je kunt je hier afvragen of bijvoorbeeld het getal 9 ('helemaal mee eens') wel een hoeveelheid is.
d Bij een gemiddelde van 2,40 zou de som 4,80 zijn en dat kan niet (de som kan 2, 3, 4, 5 of 6 zijn).
e Dat is 2,25, want bij 10 deelnemers zou de som van de ingevulde getallen dan $10 \cdot 2,25 = 22,5$ zijn en dat kan niet. (De som moet een geheel getal zijn, omdat de antwoordmogelijkheden hele getallen zijn.)
f $18 \cdot 2,63 = 47,34$ dat kan niet
 $18 \cdot 2,39 = 43,02$ dat kan
 $18 \cdot 2,17 = 39,06$ dat kan
 $18 \cdot 2,11 = 37,98$ dat kan
 $17 \cdot 1,76 = 29,92$ dat kan
 $17 \cdot 2,26 = 38,42$ dat kan niet
 $17 \cdot 1,97 = 33,49$ dat kan niet
 $17 \cdot 1,67 = 28,39$ dat kan niet
Dus in de eerste kolom van de groep met korting is 2,63 onmogelijk.
In de eerste kolom van de groep die de volle prijs betaalde zijn 2,26, 1,97 en 1,67 onmogelijk.

10.2 Causaliteit en groepen vergelijken

Bladzijde 64

- 11 a** Bij grotere lengten horen over het algemeen hogere scores.
b Ik denk het niet. Je krijgt meer kennis doordat je informatie opneemt, niet doordat je langer wordt.

Bladzijde 66

- 12 a** Voor elke categorie van discipline geldt dat het bijvoorbeeld niet zo is dat bij gezonder eten hogere scores horen.
b Als de categorie van discipline hoger is, dan zijn ook de mate van gezond eten en de scores hoger.
c De samenhang in figuur 10.13 is niet causaal, want een gemeenschappelijke factor is discipline. Dus door alleen gezonder eten aan te bieden zullen de scores niet omhoog gaan.
d Er zou een vierde factor kunnen zijn die de samenhang verklaart.
- 13 a** het warme weer
b het koude weer
c het inkomen van de gezinnen
d de grootte van de brand

Bladzijde 67

- 14 a** Nee, want niet alle gemeenschappelijke factoren zijn uitgesloten. De variabele geslacht zou bijvoorbeeld een gemeenschappelijke factor kunnen zijn die de samenhang verklaart. De jongens hadden namelijk al voor het gewichtheffen waarschijnlijk grotere bovenarmen dan de meisjes.
b Ook nu is niet uitgesloten dat er een gemeenschappelijke factor is. Het zou bijvoorbeeld kunnen dat vooral jongens zich opgaven voor het gewichtheffen.
- 15 a** Leerlingen met meer inzet zullen zowel beter leren voor toetsen als vaker de genoemde opgaven maken.
b Omdat de leerlingen aselekt zijn ingedeeld.
Alle niveaus van inzet zijn ongeveer gelijk verdeeld over de groepen, dus inzet zal niet het verschil tussen de groepen verklaren.

- c Omdat de leerlingen aselect in de groepen zijn ingedeeld zijn alle gemeenschappelijke factoren ongeveer gelijk verdeeld over de groepen. Het is daarom niet waarschijnlijk dat een andere gemeenschappelijke factor het verschil verklaart.
- d Ja, eerst kwamen de vier maanden les en daarna kwam de toets.
- e Ja, want omdat alle gemeenschappelijke factoren zijn uitgesloten is ook aan de tweede voorwaarde voldaan.

Bladzijde 68

- 16 a Van beide groepen het steekproefgemiddelde en de steekproefstandaardafwijking.
b Nee, want je weet de steekproefstandaardafwijkingen niet.

Bladzijde 69

17 a

	school		totaal
	A	B	
negatief	85	64	149
positief	43	58	101
totaal	128	122	250

b
$$\phi = \frac{85 \cdot 58 - 64 \cdot 43}{\sqrt{149 \cdot 101 \cdot 128 \cdot 122}} \approx 0,14$$

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

- c De conclusie in het voorbeeld is het beste te verdedigen. Door het omzetten van de ordinale variabele in het voorbeeld naar de nominale variabele in vraag a is informatie verloren gegaan.

Bladzijde 70

- 18 a De boxen overlappen elkaar en de medianen van de boxplots liggen buiten de box van de andere boxplot. Dus het verschil is middelmatig.
b Tussen de jaren 2008 en 2010, tussen de jaren 2008 en 2011, tussen de jaren 2008 en 2012, tussen de jaren 2008 en 2014, tussen de jaren 2008 en 2015 en tussen de jaren 2008 en 2016.
c Lisa heeft gelijk, want de andere boxen overlappen elkaar allemaal en de medianen liggen binnen de boxen van de andere boxplots.

- 19 a 54% van 108 is $0,54 \cdot 108 \approx 58$.

b

	geweldloos cum. perc.	gewelddadig cum. perc.	Vcp
mislukking	21%	62%	41%
gedeeltelijk succes	46%	75%	29%
succes	100%	100%	0%

$\max. V_{cp} = 41\%$

Omdat $\max. V_{cp} > 40$ is het verschil groot.

c

	geweldloos	gewelddadig	totaal
mislukking	23	134	157
succes	85	82	167
totaal	108	216	324

$$\phi = \frac{23 \cdot 82 - 134 \cdot 85}{\sqrt{157 \cdot 167 \cdot 108 \cdot 216}} \approx -0,38$$

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

Bladzijde 71

- 20**
- Gebruik de effectgrootte.

$$E = \frac{9,57 - 8,70}{\frac{1}{2}(0,79 + 0,83)} \approx 1,07$$

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

- 21 a** Video's over sport verliezen snel hun actualiteit. Dat is niet het geval bij video's over film.
- b** Vergeleken met muziekvideo's is
 bij sport $\max.V_{cp} = 75 - 30 = 45$ en omdat $\max.V_{cp} > 40$ is het verschil groot
 bij nieuws $\max.V_{cp} = 55 - 30 = 25$ en omdat $20 < \max.V_{cp} \leq 40$ is het verschil middelmatig
 bij film $\max.V_{cp} = 30 - 20 = 10$ en omdat $\max.V_{cp} \leq 20$ is het verschil gering.

Bladzijde 72

- 22 a**

	oorlogsjaren	vredesjaren	
koningen	735	2255	2990
koninginnen	154	442	596
	889	2697	3586

- b** oorlogsjaren koningen $\frac{735}{2990} \times 100\% \approx 24,6\%$
 oorlogsjaren koninginnen $\frac{154}{596} \times 100\% \approx 25,8\%$
 Dus koninginnen voerden relatief iets vaker oorlog.
- c** $\phi = \frac{735 \cdot 442 - 2255 \cdot 154}{\sqrt{2990 \cdot 596 \cdot 889 \cdot 2697}} \approx -0,01$
 Omdat $-0,2 \leq \phi \leq 0,2$ is het verschil gering.

- 23 a**
- Je krijgt
- $\phi = 0$
- , want
- $20 \cdot 36 - 30 \cdot 24 = 0$
- .

- b** Bij de jongens kiest $\frac{20}{50} \times 100\% = 40\%$ wiskunde A.
 Bij de meisjes kiest $\frac{24}{60} \times 100\% = 40\%$ wiskunde A.

Dit komt overeen met de conclusie van Amar.

- c** In het geval alle jongens wiskunde A kiezen en alle meisjes wiskunde B, krijg je de volgende tabel.

	jongens	meisjes	totaal
wiskunde A	50	0	50
wiskunde B	0	60	60
totaal	50	60	110

$$\text{Daarbij hoort } \phi = \frac{50 \cdot 60 - 0 \cdot 0}{\sqrt{50 \cdot 60 \cdot 50 \cdot 60}} = 1.$$

In het geval alle jongens wiskunde B kiezen en alle meisjes wiskunde A, krijg je de volgende tabel.

	jongens	meisjes	totaal
wiskunde A	0	60	60
wiskunde B	50	0	50
totaal	50	60	110

$$\text{Daarbij hoort } \phi = \frac{0 \cdot 0 - 60 \cdot 50}{\sqrt{60 \cdot 50 \cdot 50 \cdot 60}} = -1.$$

De conclusie is dat ϕ minimaal -1 en maximaal 1 is.

Bladzijde 73

- 24 a** Van de ondervraagde Chinezen geeft 65% zijn geluk een 7 of lager. Dat betekent dat 35% zijn geluk een 8 of hoger geeft. Daniek heeft dus gelijk.
- b** Gebruik de max.Vcp. Je kunt de waarden van de Vcp direct uit de tabel aflezen en vindt max.Vcp = 65 - 55 = 10.
Omdat max.Vcp ≤ 20 is het verschil gering.

c

	ongelukkig	gelukkig	totaal
China	170	830	1000
VS	220	780	1000
totaal	390	1610	2000

$$\phi = \frac{170 \cdot 780 - 830 \cdot 220}{\sqrt{1000 \cdot 1000 \cdot 390 \cdot 1610}} \approx -0,06$$

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

- 25 a** Omdat $\bar{X}_1 \geq \bar{X}_2$, is $\bar{X}_1 - \bar{X}_2$ altijd groter dan of gelijk aan 0, en de noemer is altijd positief (de standaardafwijking kan niet negatief zijn en de noemer kan niet 0 zijn).
- b** Als $\bar{X}_1$ toeneemt en de andere termen blijven gelijk, dan wordt de teller groter en daarmee neemt de effectgrootte toe.
- c** Als S_1 toeneemt en de andere termen blijven gelijk, dan wordt de noemer groter en daarmee neemt de effectgrootte af.
- d** Als $\bar{X}_1$ toeneemt en S_1 afneemt en de andere termen gelijk blijven, dan wordt de teller groter en de noemer kleiner en daarmee neemt de effectgrootte toe.
- e** Als S_1 en S_2 beide verdubbelen en $\bar{X}_1$ en $\bar{X}_2$ gelijk blijven, dan wordt de noemer twee keer zo groot en daarmee halveert de effectgrootte.

Bladzijde 74

- 26 a** Het zou kunnen dat vissen op de vorm van de luidsprekers afkomen en niet op hun geluid. Door luidsprekers zonder geluid te plaatsen kan worden uitgesloten dat de vorm een rol speelt.
- b** De interkwartielafstand is $13 - 5 = 8$ en $1,5 \cdot 8 = 12$. Anderhalf keer de interkwartielafstand boven 13 is 25 en de waarneming 34 ligt dus meer dan anderhalf keer de interkwartielafstand boven het derde kwartiel.
- c** Ze heeft niet gelijk, want zowel meer dan 18 als minder dan 5 is 25%.
- d** Met de boxplots:
groepen 1 en 2: de boxen overlappen (nog net) en de medianen liggen buiten de box van de andere boxplot, dus het verschil is middelmatig.
groepen 1 en 3: de boxen overlappen en de medianen liggen buiten de box van de andere boxplot, dus het verschil is middelmatig.
Met de effectgrootte:
groepen 1 en 2: $E = \frac{21,6 - 11,3}{\frac{1}{2}(9,8 + 9,0)} \approx 1,10$. Omdat $E > 0,8$ is het verschil groot.
groepen 1 en 3: $E = \frac{21,6 - 12,1}{\frac{1}{2}(9,8 + 8,8)} \approx 1,02$. Omdat $E > 0,8$ is het verschil groot.
- e** Aan voorwaarde 1 is voldaan, want eerst zijn de luidsprekers geplaatst en daarna is geteld hoeveel monniksvissen zich in het koraal bevonden.
Ook aan voorwaarde 2 is voldaan. Doordat de koralen aselekt zijn ingedeeld in drie groepen zijn mogelijke gemeenschappelijke factoren ongeveer gelijk verdeeld over de groepen. Deze factoren kunnen dus een verschil niet verklaren. (En ook de vorm van de luidsprekers kan niet de oorzaak zijn, omdat in de groep koralen met de dummy ook minder vissen waren.)
Dus je mag concluderen dat geluid een oorzaak is van de aanwezigheid van meer monniksvissen.

Bladzijde 75

- 27 a** De student met de hoogste score heeft ook de langste gemiddelde slaapduur (te zien in de bovenste figuur).
De student met de laagste score sliep niet gemiddeld het kortst (ook te zien in de bovenste figuur).
- b** Uitspraak I volgt uit de bovenste figuur: er zijn 6 studenten met een gemiddelde slaapduur van minstens acht uur en deze haalden allemaal tenminste 70 punten.
Uitspraak II volgt uit de onderste figuur: studenten die vaker ongeveer even lang slapen hebben een kleinere standaardafwijking van de slaapduur, en daarbij hoort een hogere score.
- c** Gebruik de effectgrootte.
- $$E = \frac{6,57}{\frac{1}{2}(13,71 + 11,01)} \approx 0,53$$
- Omdat $0,4 < E \leq 0,8$ is het verschil middelmatig.
- d** Dat kun je niet concluderen, want er zou een gemeenschappelijke factor kunnen zijn die de samenhang verklaart. Bijvoorbeeld een regelmatige dagindeling.

Bladzijde 76

- 28 a** Bijvoorbeeld wel/niet geveerd en wel/geen reactie van de koeien op de vliegen.
- b** Gebruik de effectgrootte.
Bij de zwarte koeien is $\bar{X} = 42,1 + 86,7 = 128,8$.
Bij de gestreepte koeien is $\bar{X} = 15,3 + 40,2 = 55,5$.
- $$E = \frac{128,8 - 55,5}{\frac{1}{2}(68,6 + 35,2)} \approx 1,41$$
- Omdat $E > 0,8$ is het verschil groot.
- c** Per uur is het aantal vliegen gemiddeld
bij de zwarte koeien $2 \cdot 128,8 = 257,6$ en bij de gestreepte koeien $2 \cdot 55,5 = 110$.
Een vlieg steekt elke vier minuten een keer, dus $\frac{60}{4} = 15$ steken per uur.
Het gemiddeld aantal steken per uur is
bij de zwarte koeien $15 \cdot 257,6 = 3864$ en bij de gestreepte koeien $15 \cdot 110 = 1650$.
Dus ik raad hem aan zijn koeien te verven, want 1650 is minder dan 2000 en 3864 is meer dan 2000.
- d** Om uit te sluiten dat de vliegen wegblijven vanwege de verf en niet het patroon van de verf.
- e** Alleen de Zwarte Wagyu koeien zijn onderzocht, misschien is het anders bij andere rassen. Het onderzoek vond plaats in Japan. Daar zijn misschien andere vliegen dan elders.

Bladzijde 77

- 29 a** Bij de hoge inkomens was 28% voorstander en bij de lage inkomens 69%.
- b** In het vlak linksboven staan voorstellen waarvan de hoge inkomens geen voorstander zijn en de lage inkomens wel.
- c** In het vlak linksboven is $\frac{3}{38} \times 100\% \approx 7,9\%$ van de voorstellen gerealiseerd.
- d** In het vlak rechtsonder is $\frac{10}{30} \times 100\% \approx 33,3\%$ gerealiseerd.
- Er worden verhoudingsgewijs meer voorstellen gerealiseerd waar de hoge inkomens voorstander van zijn en de lage inkomens niet, dan er voorstellen worden gerealiseerd waarvan de lage inkomens voorstander zijn en de hoge inkomens niet.
- e** Maak de volgende kruistabel.

	gerealiseerd	niet gerealiseerd	totaal
linksboven	3	35	38
rechtsonder	10	20	30
totaal	13	55	

$$\phi = \frac{3 \cdot 20 - 35 \cdot 10}{\sqrt{38 \cdot 30 \cdot 13 \cdot 55}} \approx -0,32$$

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

- f Je kunt hieruit net zo goed de conclusie trekken dat er geluisterd wordt naar mensen met hoge inkomens, want die steunen deze voorstellen ook. Het is daarom beter om te kijken naar de verschillen tussen de vlakken linksboven en rechtsonder, zoals in de vragen d en e is gedaan. En dan blijkt dat er wel verschil is in de mate waarin naar de verschillende inkomensgroepen wordt geluisterd.

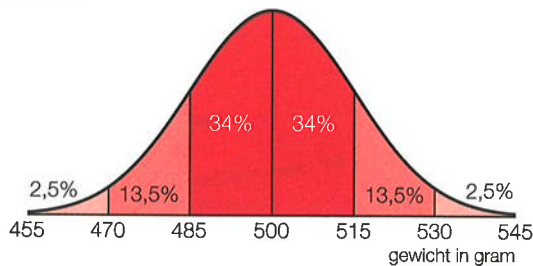
10.3 De normale verdeling en betrouwbaarheidsintervallen

Bladzijde 79

- 30 Bij figuur b, want de verdelingskromme is rechts-scheef.

Bladzijde 80

- 31 a



Zie de schets. Dus 16%.

- b 2,5%, dus dat zijn $0,025 \cdot 8000 = 200$ producten.
 c Deze hebben een gewicht tussen 470 en 530 gram.
 d Volgens de vuistregels bij de normale verdeling is bij een gemiddelde van 400 gram en een standaardafwijking van 15 gram 2,5% lichter dan 370 gram. Omdat $0,025 \cdot 3940 \approx 99$, komt een aantal van 97 ongeveer overeen met wat je op basis van een normale verdeling zou verwachten.
- 32 a $2,9 + 0,1 = 3\%$ is 190 geeft een totaal van $\frac{100}{3} \cdot 190 \approx 6300$ mannen.
 b Volgens de vuistregels zou tussen 1,72 m en 1,92 m ongeveer 68% moeten zitten. In de tabel is dat $33,2 + 35,1 = 68,3\%$. Dus weinig verschil.
 Volgens de vuistregels zou tussen 1,62 m en 2,02 m ongeveer 95% moeten zitten. In de tabel is dat $68,3 + 12,7 + 13,8 = 94,8\%$. Dus weinig verschil.
 Volgens de vuistregels zou tussen 1,52 m en 2,12 m bijna 100% moeten zitten. In de tabel is dat 99,8%. Dus ook hier weinig verschil.
- 33 a $4\sigma = 19 - 17 = 2$ mm, dus $\sigma = 0,5$ mm.
 b Hiervoor is een spreidingsdiagram geschikt.
 Het is dan nodig om per olijf de diameter en de bederftijd te weten.

Bladzijde 81

- 34 a De modus van de lengte is 165 cm (komt drie keer voor).
 Van het gewicht is er geen modus, want de gewichten 56 kg en 70 kg komen beide drie keer voor.
 b De mediane lengte (het 13^e waarnemingsgetal als op grootte is gerangschikt) is 176 cm. Bij 176 cm hoort een gewicht van 66 kg of van 72 kg. Het mediane gewicht is echter 63 kg, dus de jongen met de mediane lengte heeft niet het mediane gewicht.
 c Bij de groepen is de spreidingsbreedte achtereenvolgens $166 - 160 = 6$ cm, $172 - 168 = 4$ cm, $178 - 173 = 5$ cm, $183 - 179 = 4$ cm en $194 - 185 = 9$ cm. Dus bij de 20% langste jongens is de spreidingsbreedte het grootst.

- d 12% is 3 jongens.

De drie jongens met de grootste lengte hebben een lengte van 187 cm (die heeft een gewicht van 70 kg), 192 cm (die heeft een gewicht van 75 kg) en 194 cm (die heeft een gewicht van 70 kg).

De drie jongens met het grootste gewicht hebben een gewicht van 72 kg (die heeft een lengte van 176 cm), 74 kg (die heeft een lengte van 185 cm) en 75 kg (die heeft een lengte van 192 cm).

Dus alleen de jongen met een lengte van 192 cm en een gewicht van 75 kg hoort bij beide groepen van drie jongens. Dat is dus $\frac{1}{25} \times 100\% = 4\%$ van de jongens.

- e Ze controleert de vuistregel dat ongeveer 68% van de waarnemingsgetallen tussen $\mu - \sigma$ en $\mu + \sigma$ ligt.

In het spreidingsdiagram hebben 17 jongens een lengte die in de klasse [167, 185] valt.

Dat is dus $\frac{17}{25} \times 100\% = 68\%$.

De conclusie is dat dit overeenkomt met de vuistregel.

Bladzijde 82

- 35 a Ze maken de spreiding kleiner. Dan liggen de tonen dichter bij elkaar en zijn er dus ook meer tonen in de buurt van de hardste toon.
- b De krommen zijn niet symmetrisch.
- c Eerder een onderschatting, want de verdelingen zijn links-scheef en daarbij is het gemiddelde lager dan de modus en de mediaan.
- d Oplossen van de vergelijking $-14 + 0,067t = 0$ geeft $t \approx 209$ en hierbij hoort het jaar $1950 + 209 = 2159$.
In dat jaar klinken de tonen allemaal even hard.
- e Doordat de tonen bij *Death Magnetic* dicht bij elkaar liggen, komt de hardste toon niet ver boven het gemiddelde van -14 dBFS uit. Bij liedjes met een grote spreiding liggen tonen verder uit elkaar en ligt de hardste toon dus ook verder boven het gemiddelde van -14 dBFS.

Bladzijde 83

- 36 In de steekproef zijn 6 van de 10 bolletjes rood, dus de steekproefproportie van de rode bolletjes is $\frac{6}{10} = 0,6$.
In de populatie zijn 20 van de 34 bolletjes rood, dus de populatieproportie van de rode bolletjes is $\frac{20}{34} \approx 0,59$.

Bladzijde 84

- 37 a De populatieproportie is $\frac{50}{1000} = 0,05$.
De steekproefproportie is $\frac{4}{80} = 0,05$.
- b Nee, de steekproef is niet aselekt, omdat alleen kinderen kans hebben om in de steekproef te komen.
- 38 a Ik verwacht hoger, want geselecteerde scholen die verwachten zwak te scoren doen waarschijnlijk minder vaak mee.
- b Ik verwacht hoger, want alleen verdachte bedrijven werden onderzocht.
- 39 a De populatie is alle klimaatwetenschappers.
De steekproef is 79 klimaatwetenschappers.
- b Volgens Bram bestaat de populatie uit de wetenschappelijke artikelen waarin een mening werd gegeven.
- c Niet aselekt, want alleen de wetenschappelijke artikelen waarin een mening werd gegeven hebben kans om in de steekproef te komen.

Bladzijde 85

- 40 a** Niet alle Amerikanen die mogen stemmen gaan ook daadwerkelijk stemmen.
De populatie bestond voor de onderzoekers uit Amerikanen die gaan stemmen.
- b** Bij CNN is de proportie stemmen voor Obama gelijk aan 0,51. Dus $0,51 \cdot 1000 = 510$ gaven aan op Obama te gaan stemmen.
- c** Bij de verkiezingen zijn gegevens verzameld van alle Amerikanen die zijn gaan stemmen.
- d** Als steekproeven aselekt worden getrokken, dan wijken steekproefproporties $\hat{p}$ door toeval af van de populatieproportie p .
- e** Deze liggen in het interval $[0,51 - 2 \cdot 0,016; 0,51 + 2 \cdot 0,016] = [0,478; 0,542]$.
- f** Deze hebben een steekproefproportie van minder dan 0,478 of van meer dan 0,542.
- g** TIPP1 en TIPP2.

Bladzijde 86

- 41** De breedte is $4 \cdot 0,02 = 0,08$.

Bladzijde 87

- 42 a** De breedte van het interval is dan $4 \cdot \sqrt{\frac{0,48 \cdot 0,52}{10000}} = 0,01998...$
De breedte is dan 1,998...% en dat is minder dan 2%, dus het is nauwkeurig genoeg.
- b** $\hat{p} = 0,48$ en $n = 10000$ geeft $\sqrt{\frac{0,48 \cdot 0,52}{10000}} = 0,0049...$
 $[0,48 - 2 \cdot 0,0049...; 0,48 + 2 \cdot 0,0049...] \approx [0,470; 0,490]$
Dus het 95%-betrouwbaarheidsinterval bij het percentage is $[47,0\%; 49,0\%]$.

Bladzijde 88

- 43 a** $\left[120 - 2 \cdot \frac{36}{\sqrt{144}}, 120 + 2 \cdot \frac{36}{\sqrt{144}}\right] = [114, 126]$
- b** $\left[64,4 - 2 \cdot \frac{15,5}{\sqrt{100}}, 64,4 + 2 \cdot \frac{15,5}{\sqrt{100}}\right] = [61,3; 67,5]$
- c** $[18 - 2 \cdot 0,32; 18 + 2 \cdot 0,32] = [17,36; 18,64]$
- 44 a** $\hat{p} = 0,4$ en $n = 240$ geeft $\sqrt{\frac{0,4 \cdot 0,6}{240}} = 0,0316...$
Dus $[0,4 - 2 \cdot 0,0316...; 0,4 + 2 \cdot 0,0316...] \approx [0,337; 0,463]$.
- b** $\hat{p} = 0,48$ en $n = 380$ geeft $\sqrt{\frac{0,48 \cdot 0,52}{380}} = 0,0256...$
Dus $[0,48 - 2 \cdot 0,0256...; 0,48 + 2 \cdot 0,0256...] \approx [0,429; 0,531]$.
- c** $\hat{p} = \frac{5}{9}$ en $n = 10000$ geeft $\sqrt{\frac{\frac{5}{9} \cdot \frac{4}{9}}{10000}} = 0,0049...$
Dus $[\frac{5}{9} - 2 \cdot 0,0049...; \frac{5}{9} + 2 \cdot 0,0049...] \approx [0,546; 0,565]$.
- 45 a** Het 95%-betrouwbaarheidsinterval is
 $\left[23,4 - 2 \cdot \frac{1,9}{\sqrt{260}}, 23,4 + 2 \cdot \frac{1,9}{\sqrt{260}}\right] \approx [23,16; 23,64]$.
- b** $\hat{p} = 0,64$ en $n = 700$ geeft $\sqrt{\frac{0,64 \cdot 0,36}{700}} = 0,018...$
Het 95%-betrouwbaarheidsinterval is
 $[0,64 - 2 \cdot 0,018...; 0,64 + 2 \cdot 0,018...] \approx [0,604; 0,676]$, oftewel $[60,4\%; 67,6\%]$.
- c** Het gemiddelde is $\bar{X} = \frac{5350}{1000} = 5,35$.
Het 95%-betrouwbaarheidsinterval is $\left[5,35 - 2 \cdot \frac{3,2}{\sqrt{1000}}, 5,35 + 2 \cdot \frac{3,2}{\sqrt{1000}}\right] \approx [5,15; 5,55]$.

46 a Los op $4 \cdot \sqrt{\frac{0,3 \cdot 0,7}{n}} = 0,04$, oftewel $\sqrt{\frac{0,21}{n}} = 0,01$.

Voer in $y_1 = \sqrt{\frac{0,21}{x}}$ en $y_2 = 0,01$.

De optie snijpunt geeft $x = 2100$, dus de steekproefomvang moet minimaal 2100 zijn.

b Los op $4 \cdot \frac{4}{\sqrt{n}} = 0,8$, oftewel $\frac{16}{\sqrt{n}} = 0,8$.

Voer in $y_1 = \frac{16}{\sqrt{x}}$ en $y_2 = 0,8$.

De optie snijpunt geeft $x = 400$, dus de steekproefomvang moet minimaal 400 zijn.

47 a De breedte van het 95%-betrouwbaarheidsinterval is dus $2 \cdot 0,03 = 0,06$.

Los op $4 \cdot \sqrt{\frac{0,6 \cdot 0,4}{n}} = 0,06$, oftewel $\sqrt{\frac{0,24}{n}} = 0,015$.

Voer in $y_1 = \sqrt{\frac{0,24}{x}}$ en $y_2 = 0,015$.

De optie snijpunt geeft $x = 1066,6...$, dus de steekproefomvang moet minimaal 1067 zijn.

b De breedte van het interval is $2 \cdot 0,5 = 1$ kg.

Los op $4 \cdot \frac{2}{\sqrt{n}} = 1$, oftewel $\frac{8}{\sqrt{n}} = 1$, dus $\sqrt{n} = 8$ en $n = 64$.

De steekproefomvang moet minimaal 64 zijn.

Bladzijde 89

48 a De breedte van P is $124,1 - 112,1 = 12$.

De breedte van Q is $4 \cdot 3,3 = 13,2$.

Dus P is het meest nauwkeurig.

b Er geldt $\frac{S}{\sqrt{n}} = 3,3$, dus $\frac{66}{\sqrt{n}} = 3,3$.

Voer in $y_1 = \frac{66}{\sqrt{x}}$ en $y_2 = 3,3$.

De optie snijpunt geeft $x = 400$, dus de steekproefomvang is 400.

c Dan moet gelden dat $4 \cdot \frac{66}{\sqrt{n}} = 12$, oftewel $\frac{66}{\sqrt{n}} = 3$.

Voer in $y_1 = \frac{66}{\sqrt{x}}$ en $y_2 = 3$.

De optie snijpunt geeft $x = 484$, dus de steekproefomvang moet 484 zijn.

49 a Los op $4 \cdot \sqrt{\frac{0,1 \cdot 0,9}{n}} = 0,02$.

Voer in $y_1 = 4 \cdot \sqrt{\frac{0,1 \cdot 0,9}{x}}$ en $y_2 = 0,02$.

De optie snijpunt geeft $x = 3600$, dus hij moet minimaal 3600 mensen opnemen in zijn steekproef.

b Laura heeft gelijk. De omvang van de populatie komt niet voor in de formule van het 95%-betrouwbaarheidsinterval.

c Omdat niet iedereen geregistreerd staat is het moeilijk om een lijst op te stellen van de wereldbevolking waaruit je een aselechte steekproef kunt nemen.

Ook zijn veel mensen moeilijk te bereiken.

Bladzijde 90

50 a De populatieproportie $p = \frac{60}{261}$.

b Met de vergelijking $p - 2 \cdot \sqrt{\frac{p(1-p)}{n}} = \frac{8}{n}$, want dan is $\frac{8}{n}$ het kleinst en moet n dus het grootst zijn.

c $p - 2 \cdot \sqrt{\frac{p(1-p)}{n}} = \frac{8}{n}$ met $p = \frac{60}{261}$ geeft $\frac{60}{261} - 2 \cdot \sqrt{\frac{\frac{60}{261}(1 - \frac{60}{261})}{n}} = \frac{8}{n}$.

Voer in $y_1 = \frac{60}{261} - 2 \cdot \sqrt{\frac{\frac{60}{261}(1 - \frac{60}{261})}{x}}$ en $y_2 = \frac{8}{x}$.

De optie snijpunt geeft $x \approx 64,1$.

$p + 2 \cdot \sqrt{\frac{p(1-p)}{n}} = \frac{8}{n}$ met $p = \frac{60}{261}$ geeft $\frac{60}{261} + 2 \cdot \sqrt{\frac{\frac{60}{261}(1 - \frac{60}{261})}{n}} = \frac{8}{n}$.

Voer in $y_1 = \frac{60}{261} + 2 \cdot \sqrt{\frac{\frac{60}{261}(1 - \frac{60}{261})}{x}}$ en $y_2 = \frac{8}{x}$.

De optie snijpunt geeft $x \approx 18,9$.

Dus er kunnen minimaal 19 artikelen en maximaal 64 artikelen in de steekproef zijn geweest.

51 De stellingen II en III zijn juist.

52 a De steekproef is aselekt.

$\hat{p} = \frac{36}{119}$ en $n = 119$ geeft $\sqrt{\frac{\frac{36}{119} \cdot (1 - \frac{36}{119})}{119}} = 0,0421\dots$

Het 95%-betrouwbaarheidsinterval is

$[\frac{36}{119} - 2 \cdot 0,0421\dots; \frac{36}{119} + 2 \cdot 0,0421\dots] \approx [0,218; 0,387]$.

- b** Niet aselekt, want de steekproef is niet genomen uit de populatie, maar uit mensen die zichzelf hebben aangemeld.
c Het betreft hier geen steekproef, want de hele populatie is onderzocht.

Bladzijde 91

53 a De breedte is $4 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ met $\hat{p} = 0,45$ en $n = 1800$ en dit geeft

$4 \cdot \sqrt{\frac{0,45 \cdot 0,55}{1800}} \approx 0,047$.

In de figuur lees je af: breedte = $0,472 - 0,428 = 0,044$.

Dit komt ongeveer met elkaar overeen.

- b** De grootste peiling heeft de kleinste breedte, dus dat is de peiling van YouGov.
 De kleinste peiling heeft de grootste breedte, dus dat is de peiling van BMG.
c De breedte van het 95%-betrouwbaarheidsinterval is $0,46 - 0,42 = 0,04$ en de steekproefproportie is 0,44.

Los op $4 \cdot \sqrt{\frac{0,44 \cdot 0,56}{n}} = 0,04$.

Voer in $y_1 = 4 \cdot \sqrt{\frac{0,44 \cdot 0,56}{x}}$ en $y_2 = 0,04$.

De optie snijpunt geeft $x = 2464$, dus de steekproefomvang is 2464.

- d** Het betrouwbaarheidsinterval van BMG is heel breed. Dan kun je van tevoren niet veel zeggen over wat de uitslag gaat zijn.
- e** Bij 9 van de 12, dat is dus bij 75%.
Dat is minder dan je zou verwachten.
Een mogelijke verklaring is dat het maar 12 peilingen zijn. Als je meer peilingen zou houden, zou het percentage hoger kunnen liggen.
Het kan ook zijn dat de steekproeven niet aselekt zijn.

Bladzijde 92

54 a $\hat{p} = \frac{18}{80} = 0,225$ en $n = 80$ geeft $\sqrt{\frac{0,225 \cdot (1 - 0,225)}{80}} = 0,0466\dots$

Het 95%-betrouwbaarheidsinterval is

$$[0,225 - 2 \cdot 0,0466\dots; 0,225 + 2 \cdot 0,0466\dots] \approx [0,132; 0,318].$$

De populatieproportie is $\frac{1}{3}$ en dit ligt niet in het interval $[0,132; 0,318]$.

- b** Bij aselekte steekproeven valt de populatieproportie bij 5% van de steekproeven buiten het 95%-betrouwbaarheidsinterval. Dit zou zo'n steekproef kunnen zijn.
Ik ben het daarom niet met Lieke eens.

- 55 a** Bij een normale verdeling ligt 2,5% van de waarnemingsgetallen tussen $\mu - 3\sigma$ en $\mu - 2\sigma$.

Maar bij deze steekproef ligt 0% van de waarnemingsgetallen tussen de bijbehorende waarden $-0,017$ en $-0,005$ (want oppervlakte kan niet negatief zijn).

- b** Het 95%-betrouwbaarheidsinterval is $\left[0,019 - 2 \cdot \frac{0,012}{\sqrt{925}}; 0,019 + 2 \cdot \frac{0,012}{\sqrt{925}}\right] \approx [0,0182; 0,0198]$.

- c** De breedte van het 95%-betrouwbaarheidsinterval is $3,478 - 1,273 = 2,205$.

Dus $4 \cdot \frac{S}{\sqrt{50}} = 2,205$, oftewel $\frac{S}{\sqrt{50}} = 0,55125$.

Dit geeft $S = 0,55125 \cdot \sqrt{50} = 3,8979\dots \approx 3,898$.

- d** Los op $4 \cdot \frac{3,9}{\sqrt{n}} = 1$, oftewel $\frac{15,6}{\sqrt{n}} = 1$.

Hieruit volgt $\sqrt{n} = 15,6$ en dit geeft $n = 15,6^2 = 243,36$.

Dus het jachtgebied van Max moet minimaal 244 dagen onderzocht worden.

Bladzijde 93

- 56 a I** Lees af: het aantal katten dat een vogel mee naar huis heeft genomen is $310 + 130 + 30 + 20 + 10 = 500$.

Dus $986 - 500 = 486$ katten hebben geen vogel mee naar huis genomen.

- II** Het maximale aantal vogels dat mee naar huis is genomen is

$$5 \cdot 310 + 10 \cdot 130 + 15 \cdot 30 + 20 \cdot 20 + 25 \cdot 10 = 3950.$$

Het minimale aantal zoogdieren dat mee naar huis is genomen is

$$1 \cdot 230 + 6 \cdot 100 + \dots + 36 \cdot 10 + 41 \cdot 20 = 4518.$$

Dus er zijn meer zoogdieren dan vogels mee naar huis genomen.

- b** $\hat{p} = \frac{500}{986}$ en $n = 986$ geeft $2 \cdot \sqrt{\frac{\frac{500}{986} \cdot (1 - \frac{500}{986})}{986}} = 0,0318\dots$

Het 95%-betrouwbaarheidsinterval is $\left[\frac{500}{986} - 0,0318\dots; \frac{500}{986} + 0,0318\dots\right] \approx [0,475; 0,539]$.

- c** Per kat is het 95%-betrouwbaarheidsinterval voor het aantal mee naar huis genomen

prooien $\left[\frac{14370}{986} - 2 \cdot \frac{14,1}{\sqrt{986}}; \frac{14370}{986} + 2 \cdot \frac{14,1}{\sqrt{986}}\right] = [13,67\dots; 15,47\dots]$.

Het gevraagde 95%-betrouwbaarheidsinterval is

$$4 \cdot \frac{12}{5} \cdot 9 \cdot [13,67\dots; 15,47\dots] \approx [1182; 1337].$$

- 57** De breedte van het 95%-betrouwbaarheidsinterval voor het gemiddelde per kattenbezitter is $4 \cdot \frac{134}{\sqrt{n}}$, dus de breedte voor de 2,6 miljoen kattenbezitters is $2,6 \cdot 10^6 \cdot 4 \cdot \frac{134}{\sqrt{n}}$.

$$\text{Los op } 2,6 \cdot 10^6 \cdot 4 \cdot \frac{134}{\sqrt{n}} = 50 \cdot 10^6, \text{ oftewel } \frac{1393,6}{\sqrt{n}} = 50.$$

$$\text{Voer in } y_1 = \frac{1393,6}{\sqrt{x}} \text{ en } y_2 = 50.$$

De optie snijpunt geeft $x = 776,84\dots$

Dus minstens 777 kattenbezitters in de steekproef.

Diagnostische toets

Bladzijde 96

- 1 a** Een beoordeling op *Rotten Tomatoes* is een kwalitatieve variabele. Het aantal mensen dat positief is over een Star Wars-film per groep is een kwantitatieve variabele.
- b** Het verschil in beoordeling is het grootst bij VIII. Hierover is 90% van de recensenten en 42% van het publiek positief. Het verschil is dus $90 - 42 = 48\%$.
- c** Als je een diagonale lijn trekt in het assenstelsel, van linksonder naar rechtsboven, dan zijn alle films onder deze lijn positiever beoordeeld door recensenten dan door het publiek. Dat zijn II, III, VII en VIII.

- 2 a** kwalitatief

b

	gevangen	wild	totaal
man	24	27	51
vrouw	16	14	30
totaal	40	41	81

$$\phi = \frac{24 \cdot 14 - 27 \cdot 16}{\sqrt{51 \cdot 30 \cdot 40 \cdot 41}} \approx -0,06$$

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

- c** Met de boxplots:
De boxen overlappen elkaar en geen van beide medianen ligt buiten de box van de andere boxplot. Dus het verschil is gering.
Met de effectgrootte:
$$E = \frac{903,1 - 877,1}{\frac{1}{2}(73,3 + 55,7)} \approx 0,403$$

Omdat $0,4 < E \leq 0,8$ is het verschil middelmatig.
- d** Nee, deze conclusie mag je niet trekken, omdat je niet kunt uitsluiten dat er een gemeenschappelijke factor is. De vlinders zijn niet aselekt ingedeeld in twee groepen.

Bladzijde 97

- 3 a** Nee, want de modale klasse is 1,89-1,90 en de mediaan is het 62^e waarnemingsgetal en ligt in de klasse 1,90-1,91.
- b** De mediaan en de modus zijn niet gelijk aan elkaar. Het histogram is tweetoppig.
- c** Bij $\mu = 1,91$ en $\sigma = 0,04$ is $\mu - 2\sigma = 1,83$. Dat betekent dat 2,5% van de waarnemingsgetallen kleiner dan 1,83 zou zijn als er sprake was van een normale verdeling.
In figuur 10.35 lees je af dat 3 waarnemingsgetallen kleiner zijn dan 1,83. Dat is $\frac{3}{123} \times 100\% = 2,439\dots\%$. De percentages komen dus ongeveer overeen.

4 a $\hat{p} = \frac{55}{220} = 0,25$ en $n = 220$ geeft $\sqrt{\frac{0,25 \cdot 0,75}{220}} = 0,0291\dots$

$[0,25 - 2 \cdot 0,0291\dots; 0,25 + 2 \cdot 0,0291\dots] \approx [0,192; 0,308]$.

Dus $[19,2\%; 30,8\%]$.

b Nederlanders zonder longziekte hadden geen kans om in de steekproef te komen.

5 a $\bar{X} = \frac{284}{50} = 5,68$, $S = 2,2$ en $n = 50$

$\left[5,68 - 2 \cdot \frac{2,2}{\sqrt{50}}; 5,68 + 2 \cdot \frac{2,2}{\sqrt{50}} \right] \approx [5,06; 6,30]$.

b Los op $4 \cdot \sqrt{\frac{0,5 \cdot 0,5}{n}} = 0,04$, oftewel $\sqrt{\frac{0,25}{n}} = 0,01$.

Voer in $y_1 = \sqrt{\frac{0,25}{x}}$ en $y_2 = 0,01$.

De optie snijpunt geeft $x = 2500$, dus het aantal personen in de steekproef moet minimaal 2500 zijn.