

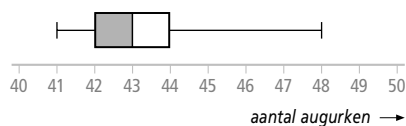
Hoofdstuk 6 - Statistische verwerking

bladzijde 160

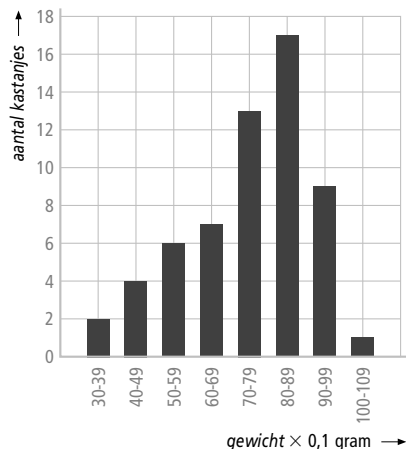
- V-1a** Het gemiddelde salaris is
- $$\frac{1 \times 15000 + 2 \times 10000 + 2 \times 8000 + 3 \times 4000 + 6 \times 3000 + 11 \times 2000 + 1 \times 1000}{1 + 2 + 2 + 3 + 6 + 11 + 1} =$$
- $$\frac{104000}{26} = 4000 \text{ euro.}$$
- b** Van 26 getallen is de mediaan het gemiddelde van het 13^e en 14^e getal (nadat de getallen zijn gerangschikt naar opklimmende grootte). De mediaan is hier dus
- $$\frac{3000 + 3000}{2} = 3000 \text{ euro.}$$
- c** Het modale salaris is het salaris met de hoogste frequentie, dat is dus € 2000.
- V-2a** De som van de frequenties is het aantal leerlingen in een klas. Voor klas 2A is dat $0 + 0 + 1 + 3 + 8 + 2 + 4 + 3 + 3 + 0 = 24$ leerlingen en het blijkt dat ook klas 2B $0 + 3 + 4 + 4 + 0 + 10 + 2 + 1 + 1 + 1 = 24$ leerlingen heeft.
- b** De modus is het cijfer met de hoogste frequentie, dat is 5. De mediaan is het gemiddelde van het 12^e en 13^e cijfer, dus $\frac{5+6}{2} = 5,5$. Het gemiddelde cijfer voor klas 2A is
- $$\frac{1 \times 3 + 3 \times 4 + 8 \times 5 + 2 \times 6 + 4 \times 7 + 3 \times 8 + 3 \times 9}{24} = \frac{146}{24} \approx 6,1.$$
- c** Voor klas 2B is de modus het cijfer 6, de mediaan $\frac{6+6}{2} = 6$ en het gemiddelde is
- $$\frac{116}{24} \approx 4,8.$$
- d** Klas 2B doet het slechter en dat zie je vooral aan het gemiddelde.
- e** De modus en de mediaan veranderen niet. In de nieuwe situatie is 6 nog steeds het meest voorkomende cijfer en zijn het 12^e en 13^e cijfer nog steeds zessen.
- f** Mediaan: als alle cijfers één punt worden verhoogd, dan wordt het middelste waarnemingsgetal ook één punt hoger, dus de mediaan ook. Modus: als alle cijfers één punt worden verhoogd, dan is ook de waarneming met de hoogste frequentie één punt hoger en dus ook de modus. Gemiddelde: als alle 24 cijfers één punt worden verhoogd, wordt de som van alle cijfers 24 hoger. Delen door 24 geeft een gemiddelde dat precies één hoger is.
- V-3a** De mediaan is volgens de definitie het midden van de waarnemingsgetallen. Er zijn in het algemeen 50% waarnemingen links en 50% waarnemingen rechts van de mediaan.
- b** Neem de waarnemingsgetallen 1, 8, 9, 10. Deze reeks heeft het gemiddelde 7, zodat 25% van de waarnemingen links en 75% van de waarnemingen rechts van het gemiddelde ligt.

bladzijde 161

- V-4** De kleinste waarde is 41 en de grootste 48. De mediaan is het gemiddelde van het aantal augurken in het 125^e en 126^e blik, dus $\frac{43+43}{2} = 43$. Q_1 is het aantal augurken in de middelste van de eerste 125 blikken. Je moet dus het 63^e blik hebben en $Q_1 = 42$. Q_3 is het aantal augurken in de middelste van de rechter 125 blikken. Het gaat dus om het 63^e blik geteld vanaf rechts en dus $Q_3 = 44$.



- V-5** Alhoewel er een zesdeklasser is geweest die meer tijd in biologie heeft gestoken dan de tijd die wie dan ook in aardrijkskunde heeft gestoken, is er over het algemeen toch meer tijd besteed aan aardrijkskunde. Bijvoorbeeld heeft 75% van de leerlingen minstens 40 uur besteed aan aardrijkskunde, terwijl dat percentage voor biologie slechts op 50% ligt.
- V-6a** Er zijn drie kastanjes van 8,0 gram.
- b** De zwaarte is 10,1 gram, de lichtste 3,2 gram, een verschil van 6,9 gram.
- c** De waarneming met de hoogste frequentie is 7,7 (komt $4 \times$ voor), de modus is dus 7,7 gram. Er zijn 59 kastanjes. De mediaan is dan het gewicht van de 30^e kastanje. In het steeldiagram zijn de gewichten al naar grootte gesorteerd, dus is gemakkelijk af te lezen dat de mediaan 7,7 gram is. Het bepalen van het gemiddelde is iets bewerkelijker. Bij het diagram kun je de som van de gewichten sneller vinden door handig te tellen: de 6 gewichten in het rijtje 5 | 1 5 5 6 7 9 bijvoorbeeld resulteren in $5 \times 6 + 0,1 \times (1 + 5 + 5 + 6 + 7 + 9) = 33,3$ gram. Het hele diagram resulteert zo in een gewicht van $6,5 + 17,4 + 33,3 + 47,0 + 96,2 + 144,8 + 85,7 + 10,1 = 441,0$ gram en het gemiddelde wordt dan $\frac{441}{59} \approx 7,47$ gram.
- d** Met VU-Stat kun je het steelbladdiagram invoeren en als staafdiagram weergeven.



Voordelen van het steelbladdiagram ten opzichte van het staafdiagram:

- Er is geen informatieverlies, terwijl dat bij het maken van een staafdiagram vaak wel het geval is
- Je kunt een steelbladdiagram ook zien als een soort horizontaal georiënteerd staafdiagram. De overzichtelijkheid is dus groter dan bij zo maar een tabel.
- Een steelbladdiagram is niet meer handig als het aantal gegevens te groot wordt.

bladzijde 162

- 1a** In 1995 werden er 190 500 kinderen geboren. Per 1000 vrouwen van 15-44 jaar werden er 55,4 kinderen geboren. Bij benadering moeten er in 1955 dus

$$\frac{190500}{55,4} \times 1000 = 3,4 \text{ miljoen vrouwen van 15-44 jaar zijn geweest.}$$

- b Het aantal geboren kinderen stijgt, maar dat kan ook gebeuren bij een afnemende bevolking. Door berekening kun je bepalen wat er hier speelt. In 1995 werden 190 500 kinderen geboren en per 1000 inwoners 12,3. Dus waren er ongeveer $\frac{190500}{12,3} \times 1000 = 15,5$ miljoen inwoners. Op vergelijkbare manier kun je inzien dat er in 2000 ongeveer $\frac{205300}{12,9} \times 1000 = 15,9$ miljoen inwoners waren. Het inwonertal is dus in de periode 1995-2000 iets toegenomen.
- c Er zullen in 2010 naar verwachting $\frac{179500}{56,2} \times 1000 \approx 3,2$ miljoen vrouwen zijn van 15-44 jaar.

2a Voer in je rekenmachine in $L_3 = L_2 / \text{sum}(L_2) * 100$ geeft:

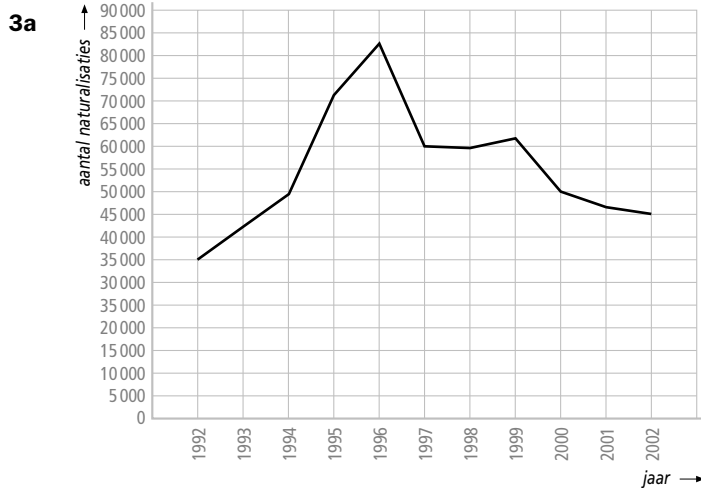
L1	L2	L3	3
12	312	25.263	
13	420	34.008	
14	351	28.421	
15	152	12.308	
-----	-----	-----	
<hr/>			
L3(1)=25.26315789...			

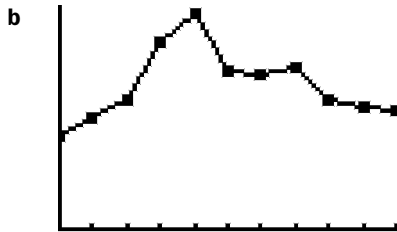
b Behalve de data voor 2000 moet je ook opnieuw $L_3 = L_2 / \text{sum}(L_2) * 100$ invoeren.

L1	L2	L3	3
12	296	25.263	
13	375	33.098	
14	319	28.155	
15	143	12.621	
-----	-----	-----	
<hr/>			
L3(1)=26.12533097...			

- c In 1990 (152 > 143)
 d In 2000 (12,621 > 12,308)

bladzijde 163

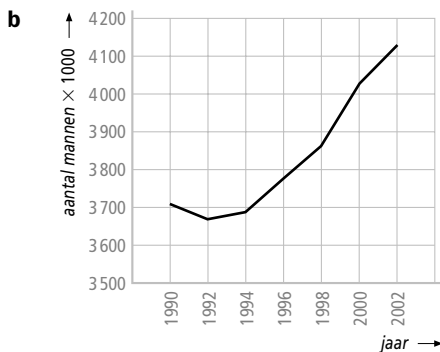




- c Tussen 1996 en 1997. De grafiek loopt daar het steilst naar beneden.
 d Er is geen informatie per maand aanwezig. Je kunt niet simpelweg het aantal naturalisaties in het jaar 1995 door 12 delen.

4a Voer in 3 lijsten: L_1 : de jaartallen, L_2 : de gegevens over mannelijke autobezitters, L_3 : de gegevens over vrouwelijke autobezitters. Dan is $L_4 = L_2 + L_3$ voor de totale aantallen autobezitters.

L1	L2	L3	1	L2	L3	L4	4
1990	3710	1486		3710	1486	5196	
1992	3670	1626		3670	1626	5296	
1994	3689	1869		3689	1869	5558	
1996	3779	1962		3779	1962	5741	
1998	3864	2186		3864	2186	6050	
2000	4027	2452		4027	2452	6479	
2002	4131	2673		4131	2673	6804	
$L_1(0) = 1990$				$L_4(0) = 5196$			



Het aantal mannen met een auto steeg het sterkst van 1998 tot 2000

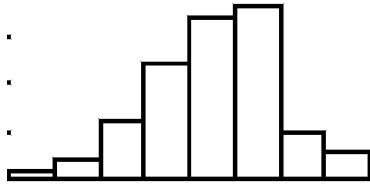
- c Definieer bijvoorbeeld $L_5 = L_2/L_4 * 100$ voor de mannelijke autobezitters en $L_6 = L_3/L_4 * 100$ voor de vrouwelijke autobezitters.

jaar	1990	1992	1994	1996	1998	2000	2002
% mannen met auto	71,4	69,3	66,4	65,8	63,9	62,2	60,7
jaar	1990	1992	1994	1996	1998	2000	2002
% vrouwen met auto	28,6	30,7	33,6	34,2	36,1	37,8	39,3

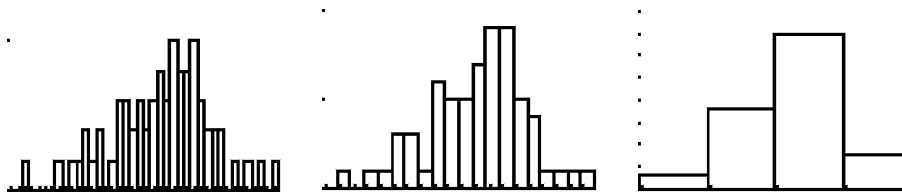
- d Absoluut gezien is in de loop van 1990-2002 het aantal mannelijke autobezitters gestegen en dat geldt ook voor de vrouwelijke autobezitters ($4134 > 3710$ en $2673 > 1486$).
 e Relatief gezien is het aantal vrouwen met een auto in de periode van 1990-2002 toegenomen en is het aantal mannelijke autobezitters dus afgenomen ($39,3 > 28,6$ en $60,7 < 71,4$).

bladzijde 164

5b Als linkergrens van de eerste klasse is 139,5 gekozen.



c Als linkergrens van de eerste klasse is steeds 139,5 genomen. De klassenbreedtes zijn resp. 1 cm, 2 cm en 10 cm.



Bij kleine klassenbreedtes is zo'n staafdiagram wel erg gedetailleerd en daardoor ook grillig, terwijl bij grotere klassenbreedtes wat meer informatie verloren gaat maar een globaal beeld ontstaat. In het extreme geval van een klassenbreedte van 40 cm is er maar één staaf is alle detailinformatie verloren gegaan en dat gaat natuurlijk veel te ver. In het voorbeeld biedt een klassenbreedte van 5 cm een goed compromis.

6a,b,c Kies de klassen en turf hoeveel waarnemingen in elke klasse vallen.

<i>klasse</i>	<i>frequentie</i>	<i>klassenmiddens</i>
$[139,5; 144,5)$	1	142
$[144,5; 149,5)$	2	147
$[149,5; 154,5)$	6	152
$[154,5; 159,5)$	12	157
$[159,5; 164,5)$	17	162
$[164,5; 169,5)$	18	167
$[169,5; 174,5)$	5	172
$[174,5; 179,5)$	3	177

d Je schat het gemiddelde nu met behulp van de klassenmiddens.

$$\frac{1 \times 142 + 2 \times 147 + 6 \times 152 + 12 \times 157 + 17 \times 162 + 18 \times 167 + 5 \times 172 + 3 \times 177}{64} =$$

$$\frac{10383}{64} \approx 162,2 \text{ cm.}$$

e Voer in je rekenmachine in lijst L_1 de oorspronkelijke data in. Neem vervolgens $\text{sum}(L_1)$ en op het scherm verschijnt de som, namelijk 10382. Delen door 64 levert ook hier het gemiddelde van ongeveer 162,2.

f,g Zie uitwerkingen bij opdracht 5c.

bladzijde 165

- 7a** Het aantal verkochte kaartjes is steeds een geheel getal. In de klasse 70-79 zitten de verkochte aantallen 70, 71, 72, 73, 74, 75, 76, 77, 78 en 79. De volgende klasse is 80-89, enzovoort.
- b** Bij gewichten heb je te maken met afgeronde getallen. En er wordt hier afgerond op gehele getallen: een gewicht $\geq 69,5$ maar $< 70,5$ wordt afgerond op 70, enzovoort. De feitelijke klasse is hier $[69,5; 79,5)$ en de volgende klasse is $[79,5; 89,5)$.
- c** Bij leeftijden rond je getallen niet af. Iemand van 72,9 jaar is nog steeds 72 jaar en niet 73. In feite wordt het getal bij de komma afgekapt. De klasse 70-79 is dus eigenlijk de klasse $[70; 80)$ en de volgende klasse is $[80; 90)$.
- d** Bij de verkochte kaartjes is de klassenbreedte 9 en het klassenmidden $\frac{70+79}{2} = 74,5$.

Bij de gewichten is de klassenbreedte 10 en het klassenmidden $\frac{69,5+79,5}{2} = 74,5$.

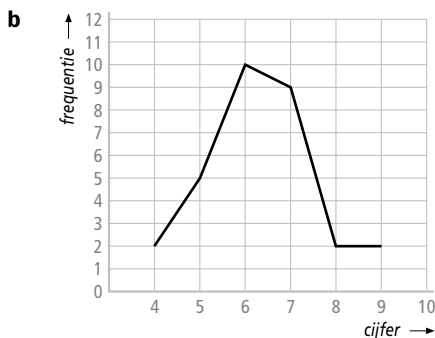
Bij de leeftijden is de klassenbreedte 10 en het klassenmidden $\frac{70+80}{2} = 75$.

- 8a** Na afronding op gehele getallen krijg je de tabel:

<i>cijfer</i>	<i>frequentie</i>
4	2
5	5
6	10
7	9
8	2
9	2

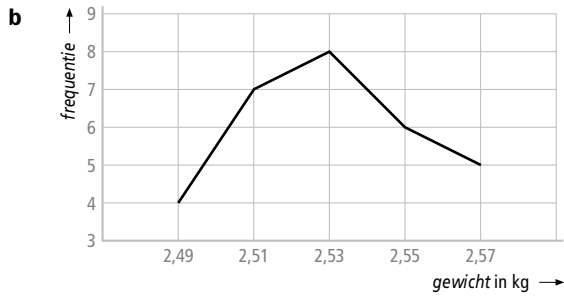
De frequentietabel wordt dan:

In de klasse 4 vallen de cijfers uit het interval $[3,5; 4,5)$ met klassenmidden 4. De klassenbreedte is $4,5 - 3,5 = 1$, wat ook geldt voor de andere klassen.



9a

<i>klassenmidden</i>	<i>frequentie</i>
2,49	4
2,51	7
2,53	8
2,55	6
2,57	5

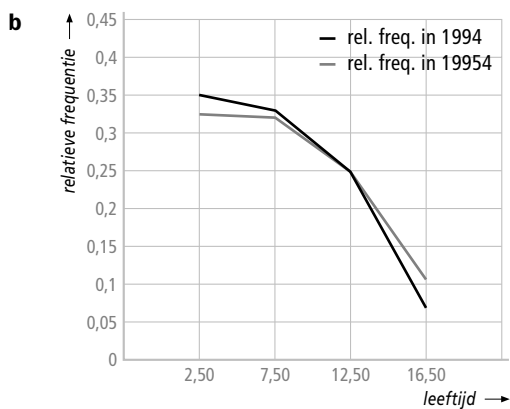


- c** De verdeling is redelijk symmetrisch en dus zal het gemiddelde in de buurt liggen van het klassenmidden van de middelste klasse. De schatting is dus 2,53 kg.
- d** Op grond van de klassenmiddens kunnen we het gemiddelde nog preciezer schatten:

$$\frac{(4 \times 2,49 + 7 \times 2,51 + 8 \times 2,53 + 6 \times 2,55 + 5 \times 2,57)}{30} = \frac{75,92}{30} \approx 2,53 \text{ kg.}$$
- e** De $4 + 7 = 11$ zakken die in de eerste twee klassen tuishoren hebben in ieder geval een gewicht dat minder is dan 2,53 kg. In de middelste klasse, namelijk $[2,52; 2,54)$ horen 8 zakken aardappels. Hoe de 8 zakken precies over deze klasse zijn verdeeld, weet je niet, maar je neemt aan dat die verdeling gelijkmatig is. Dus bijvoorbeeld 4 zakken lichter en 4 zakken zwaarder dan 2,53 kg. In totaal zijn er dus dan al $11 + 4 = 15$ zakken lichter dan 2,53 kg.

- 10a** In de leeftijdsklasse 0-4 vallen alle kinderen van 0 tot 5 jaar. Het klassenmidden is 2,5 jaar. De relatieve frequentie in 1994 voor kinderen van 0-4 jaar is $\frac{157}{452} \approx 0,347$.

klassenmidden	1994	1995
2,5	0,347	0,325
7,5	0,330	0,321
12,5	0,252	0,247
16,5	0,071	0,107



De grafiek van de relatieve frequenties in 1994 daalt het snelst. In 1994 werd relatief vaker bij jongere kinderen één of beide ouders uit de ouderlijke macht ontzet.

bladzijde 166

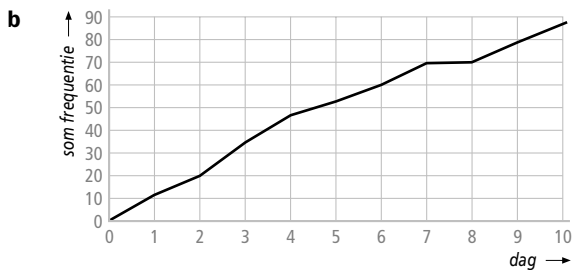
- 11a** Een score die minimaal 3,5 is en lager dan 4,5 wordt steeds afgerond op cijfer 4. De grenzen van deze klasse zijn dus 3,5 en 4,5.
- b** Het gaat om de leerlingen die een 3 of een 4 hebben gescoord. Dat zijn er 5.

c De leerlingen die hoogstens eindcijfer 4 hebben gehaald (5 stuks, zie onderdeel b) aangevuld met de leerlingen met eindcijfer 5 (9 stuks). Dus $5 + 9 = 14$.

eindcijfer	frequentie	frequentie voor hoogstens dat eindcijfer
3	2	2
4	3	5
5	9	14
6	13	27
7	7	34
8	4	38
9	2	40

12a

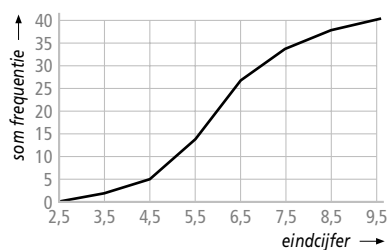
dag	1	2	3	4	5	6	7	8	9	10
aantal	12	8	15	12	6	7	10	0	9	8
totaal aantal	12	20	35	47	53	60	70	70	79	87



c Het steilst loopt de grafiek van dag 2 naar dag 3. Op dag 3 heeft hij dus het grootste aantal kistjes bollen gepeld.

bladzijde 167

13a De klassen bij de eindcijfers zijn $[2,5; 3,5)$, $[3,5; 4,5)$, $[4,5; 5,5)$, enzovoort. De somfrequenties zijn boven de rechter klassengrenzen uitgezet.



b Hoogstens een eindcijfer 6 betekent minder dan beginscore 6,5. Lees af bij 6,5 dat het aantal leerlingen met eindcijfer van hoogstens een 6 gelijk is aan 27. Het aantal leerlingen met eindcijfer hoger dan 6 is dan $40 - 27 = 13$.

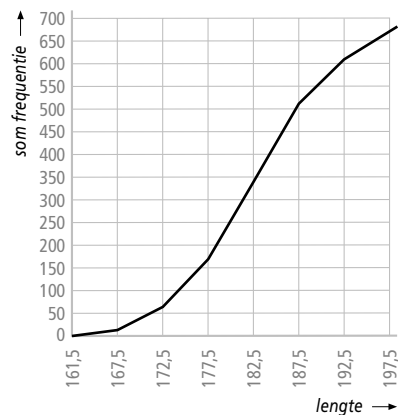
14a Voer in de klassenmiddens (lijst L_1) en bijbehorende frequenties (lijst L_2) en definieer $L_3 = \text{cumSum}(L_2)$.

L_1	L_2	L_3
164,5	13	13
170	51	64
175	106	170
180	169	339
185	172	511
190	96	607
195	63	670

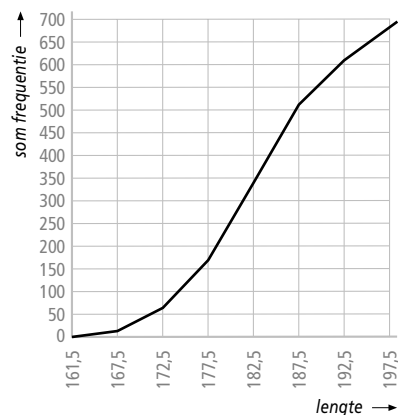
$L_3 = \{13, 64, 170, 339, 511, 607, 670\}$

Je kunt de tabel van het boek op de volgende manier aanvullen:

lengte in cm	klasse	aantal mannen	cumulatieve frequentie
162-167	$[161,5; 167,5)$	13	13
168-172	$[167,5; 172,5)$	51	64
173-177	$[172,5; 177,5)$	106	170
178-182	$[177,5; 182,5)$	169	339
183-187	$[182,5; 187,5)$	172	511
188-192	$[187,5; 192,5)$	96	607
193-197	$[192,5; 197,5)$	63	670



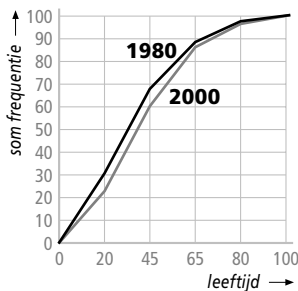
- b Het steilste gedeelte bevindt zich tussen 182,5 en 187,5 cm. Het aantal mannen met een lengte tussen deze grenzen is het grootst.
- c De 14 mannen van 196 cm vallen in de laatste klasse, zodat het aantal in die klasse van 63 naar 77 gaat. Bijbehorende cumulatieve frequentie verandert van 670 naar 684. De andere cumulatieve frequenties veranderen niet. De grafiek zal vrijwel identiek zijn aan die bij onderdeel a. Alleen van 192,5 tot 197,5 gaat de grafiek ineens wat steiler lopen.



- d De 20 mannen van 165 cm vallen in de klasse $[161,5; 167,5)$ zodat bijbehorende aantal van 13 naar 33 gaat. Tussen 161,5 en 167,5 zal de grafiek dus iets steiler lopen. De rest van de grafiek zal 20 hoger komen te liggen omdat alle cumulatieve frequenties groter 20 zijn. Tussen 161,5 en 167,5 zal de grafiek iets steiler lopen

15a Bij leeftijden rond je de jaren niet af. Iemand van 19 jaar kan bijvoorbeeld op een dag na wel 20 jaar zijn. In de leeftijdsklasse 0 – 19 zitten dus alle leeftijden uit het interval $[0; 20)$, zodat de rechter klassengrens voor die klasse 20 is.

klasse	rel.freq. in 1980	cum.rel.freq. in 1980	rel.freq. in 2000	cum.rel.freq. in 2000
0-19	31,5	31,5	23,4	23,4
20-44	37,2	68,7	38,0	61,4
45-64	19,9	88,6	25,1	86,5
65-79	9,3	97,9	10,5	97,0
≥ 80	2,2	100,1	3,0	100,0



- c** Voor 1980 hoort bij 31,5% de leeftijd van 20 jaar en bij 68,7% die van 45 jaar. De mediaan ligt bij 50% van de verdeling. Door interpolatie vind je dat bij de mediaan een leeftijd van ongeveer 32,4 hoort. Voor 2000 vind je op vergelijkbare manier 37,5 jaar als mediane leeftijd.
- d** Als 25% van de bevolking ouder is dan de gezochte leeftijd, dan is 75% jonger. Je kunt de gezochte leeftijd dus aflezen bij 75% in de somfrequentiepolygoon. Je vindt ongeveer 53 jaar voor 1980 en 56 jaar voor 2000.

bladzijde 168

- 16a** Bij de stippellijn lees je af dat het gemiddelde gewicht ongeveer 142 lb is, dat is $142 \times 0,454 \approx 64,5$ kg.
- b** Lees uit het staafdiagram af de cumulatieve relatieve frequenties behorende bij de klassen $[90; 100)$ en $[100; 110)$. Dit geeft respectievelijk ca. 2% en ca. 7%. Ongeveer $2\% + 7\% = 9\%$ van de Amerikaanse vrouwen, dus een kleine 10% is lichter dan 110 lb.
- c** Onder het gemiddelde gewicht vind je de eerste vijf staven en een deel van de zesde staaf. De eerste vijf staven bij elkaar opgeteld geven een percentage van ongeveer $2\% + 7\% + 13\% + 16\% + 16\% = 54\%$. De gewichten 140-142 lb vormen het $\frac{1}{5}$ deel van de gewichten binnen de staaf 140-150. Deze staaf bevat ongeveer 14,5 % van de vrouwen, dus naar verwachting zal $\frac{1}{5} \times 14,5\% = 2,9\%$ van de vrouwen in de zesde klasse en onder het gemiddelde gewicht vallen. Het te schatten percentage is $54\% + 2,9\% = 56,9\%$.
- d** De mediaan is het gewicht waaronder 50% van de waarneming ligt. Van links af beginnend kom je t/m de vierde klasse op ongeveer $2\% + 7\% + 13\% + 16\% = 38\%$. De vijfde klasse bevat 16%. Met die klasse erbij zit je al op 54%. De mediaan moet dus in de klasse 130-140 zitten. Door interpolatie vind je als mediaan $130 + \frac{12}{16} \times (140 - 130) = 137,5$ lb. Dat is $137,5 \times 0,454 \approx 62,4$ kg.
- e** De mediaan is kleiner dan het gemiddelde omdat er aan de rechterkant van de verdeling een aantal zwaargewichten is die het gemiddelde naar boven trekt.

- 17a** De mediaan is bij boxplot B gelijk aan 5,5, dus heeft 50% van de vierde klas een onvoldoende gehaald voor de toets statistiek.
- b** De mediaan in boxplot A is gelijk aan 6. Dus 50% van de vierdeklassers heeft voor de toets analyse hoger dan een 6 gescoord. Tussen de 5,5 en 6 zitten ook nog leerlingen die een voldoende hebben gehaald. Het percentage voldoende is dus hoger dan 50%.
- 18a** Dit betekent dat de werkloosheid onder mannen tussen februari 1986 en februari 1987 met $38 \times 1000 = 38000$ afnam.
- b** Die is met ongeveer $3 \times 1000 = 3000$ gestegen.
- c** Aan het eind van 1987 waren er minder mannen werkloos dan aan het eind van 1986. Dat zie je aan de negatieve waarde bij december.
- d** Als de werkloosheid in 1987 onder vrouwen gelijk was aan die in 1986 dan zijn er in 1987 evenveel vrouwen werkend als werkloos geworden. De som van de staven naar beneden (werkend geworden) moet dus gelijk zijn aan de som van de staven naar boven (werkloos geworden).

bladzijde 169

- 19a** Bij 2400 uren lees je ongeveer 40% af, dus ongeveer 40% van de lampen had een levensduur van 2400 uur of korter.
- b** Je moet kijken bij 50%. Per definitie ligt 50% van de data onder de mediaan. De mediaan is ongeveer 2800.
- c** Het eerste kwartiel Q_1 is de mediaan van de onderste 50% en ligt dus bij 25%. Je leest af ongeveer 1550 branduren. Het derde kwartiel Q_3 vind je bij 75%. Je leest af ongeveer 3250 branduren.
- d** Om een boxplot te tekenen, heb je behalve de mediaan, Q_1 en Q_3 ook nog de waarden van de laagste en de hoogste waarneming nodig. Ook die lees je af uit de polygoon: 500 resp. 4000 branduren.



- 20a** Uit de grafiek blijkt dat jaarlijks het aantal fietsendiefstallen met gemiddeld 1,8% toenam gedurende de periode 1980-1992. Als dat percentage ook geldt voor de periode 1992-1993 krijg je voor 1993 het aantal van $746\,700 \times 1,018 \approx 760\,100$ diefstallen. Het jaar 1991 was een jaar eerder dus er waren toen $\frac{746\,700}{1,018} \approx 733\,500$ fietsendiefstallen.
- b** Elk jaar namen de diefstallen uit woonhuizen met 8,1% toe, dus de groeifactor was 1,081 per jaar. Per twee jaar krijg je dan de groeifactor van $1,081 \times 1,081 \approx 1,169$ en een procentuele toename van ongeveer 16,9%.
- c** Het aantal diefstallen uit de auto nam jaarlijks met 3,1% af. De jaarlijkse groeifactor voor dit soort diefstallen is 0,969. Over 12 jaar betekent dat de groeifactor $0,969^{12} \approx 0,685$ en een procentuele afname van ongeveer 31,5%.

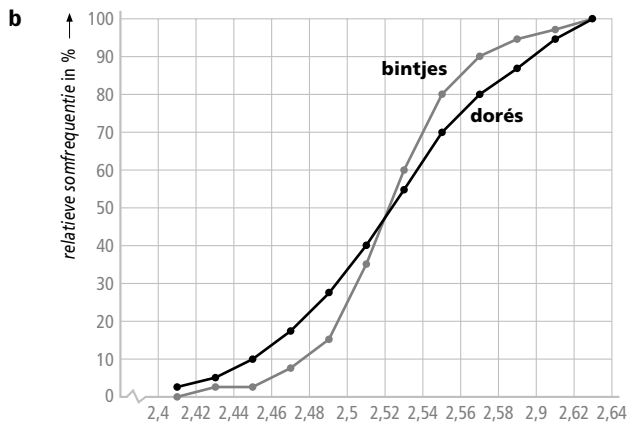
- 21a** Het antwoord moet zijn: 173 000 huishoudens. Dit is als volgt in te zien. In 1981 waren er in totaal 313 000 huishoudens waarvan 225 000 doorgingen naar 1982, er 68 000 definitief afvloeiden en er 20 000 pas in 19983 terugkeerden. In 1982 waren er in totaal 468 000 huishoudens die een uitkering kregen. Van de 22 5000 huishoudens uit 1981stroomden er 173 000 door naar 1983 en vloeiden er 52 000 definitief af.
- b** De som is 567 en je komt dus 1 te kort. Door de afronding op duizendtallen die steeds wordt gehanteerd kan de optelling een klein verschil geven.
- c** Dit moet $468 - 225 = 243$ zijn.
- d** Het totaal aantal huishoudens uit de figuur is $313 + 243 + 216 = 772$ (de doorstomers tellen maar één keer mee). De huishoudens met één keer de eenmalige uitkering zijn te vinden in de stromen met 68, 85 en 216. Het gevraagde percentage is $\frac{68 + 85 + 216}{772} \times 100\% \approx 47,8\%$.
- e** Er komen dan zoveel stromen dat dit te zeer ten koste gaat van de overzichtelijkheid.

bladzijde 170

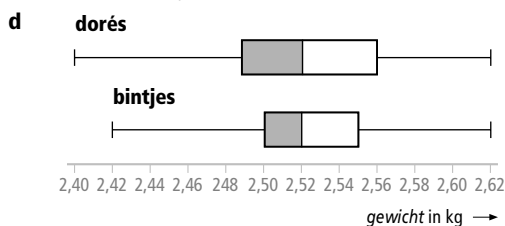
- 22a** Nee. In de ene klas kan iedereen een 7,3 hebben gehaald, terwijl in de andere klas alleen extreem hoge en extreem lage cijfers voorkomen. De eerste klas is homogeen qua niveau, de tweede klas extreem heterogeen.
- b** De spreiding van de cijfers in de klassen liep nogal uiteen. In klas 3a lagen de cijfers redelijk dicht bij elkaar in de buurt terwijl in klas 3b de cijfers voor een groot deel ver uit elkaar lagen.
- 23a** De winnaar is degene die het eerst aankomt, dus de laagste tijd scoort. Je leest ongeveer 135 minuten af.
- b** De langzaamste loper doet er 235 minuten over. Het verschil tussen snelste en langzaamste loper is 100 minuten of 1 uur en 40 minuten.
- c** Twee uur en 40 minuten komt overeen met 160 minuten. In de boxplot zie je dat 160 het eerste kwartiel is. 25% van de 200 lopers finisht dus binnen die tijd, dus 50 lopers.
- d** De 50^e tijd van de 200 tijden is ongeveer het eerste kwartiel, de 100^e tijd is ongeveer de mediaan. Het verschil tussen bijbehorende tijden is dan $170 - 160 = 10$ minuten.
- e** De tijd van de als 150^e binnenkomende loper is ongeveer het derde kwartiel. De laatste loper doet er 235 minuten over. Het verschil in tijd tussen deze twee lopers is ongeveer 37 minuten.
- 24a** De kleinste waarde is 25 en de grootste 79. De spreidingsbreedte is $79 - 25 = 54$.
- b** Er zijn 30 scores. Alle scores zijn gerangschikt naar grootte. Je moet dus het gemiddelde nemen tussen de 15^e en 16^e waarneming. De mediaan is $\frac{44 + 45}{2} = 44,5$.
- c** De 8^e score geeft het eerste kwartiel Q_1 met 38 en de 23^e score geeft het derde kwartiel Q_3 met 48. De kwartielafstand is dus $48 - 38 = 10$.

bladzijde 171

gewichtsklasse in kg	aantal Bintjes	cum. freq. Bintjes	cum. rel. freq. Bintjes	aantal Dorés	cum. freq. Dorés	cum. rel. freq. Dorés
(2,39; 2,41]	0	0	0%	1	1	2,5%
(2,41; 2,43]	1	1	2,5%	1	2	5%
(2,43; 2,45]	0	1	2,5%	2	4	10%
(2,45; 2,47]	2	3	7,5%	3	7	17,5%
(2,47; 2,49]	3	6	15%	4	11	27,5%
(2,49; 2,51]	8	14	35%	5	16	40%
(2,51; 2,53]	10	24	60%	6	22	55%
(2,53; 2,55]	8	32	80%	5	27	67,5%
(2,55; 2,57]	4	36	90%	5	32	80%
(2,57; 2,59]	2	38	95%	3	35	87,5%
(2,59; 2,61]	1	39	97,5%	3	38	95%
(2,61; 2,63]	1	40	100%	2	40	100%



c De modale klasse is de klasse waarin de somfrequentiepolygoon het sterkst stijgt. In beide gevallen is dat (2,51; 2,53]. De spreidingsbreedte bij een klassenindeling is het verschil tussen de klassenmiddens van de hoogste en de laagste klasse. Bij de Bintjes zijn deze klassenmiddens 2,62 kg en 2,42 kg, dus is de spreidingsbreedte $2,62 - 2,42 = 0,20$ kg. Bij de Dorés vind je 2,62 kg en 2,40 kg, dus de spreidingsbreedte $2,62 - 2,40 = 0,22$ kg. De mediaan is het gewicht dat hoort bij 50% van de relatieve cumulatieve frequentie. Aflezen geeft afgerond 2,52 kg voor beide aardappelsoorten. De kwartielen zijn de gewichten die horen bij 25% en 75% van de relatieve somfrequentie. Aflezen geeft bij de Bintjes $Q_1 \approx 2,50$ kg en dus is de kwartielafstand gelijk aan $Q_3 - Q_1 = 0,05$ kg. Bij de Dorés vind je $Q_1 \approx 2,49$ kg en $Q_3 \approx 2,56$ kg en is $Q_3 - Q_1 = 0,07$ kg.

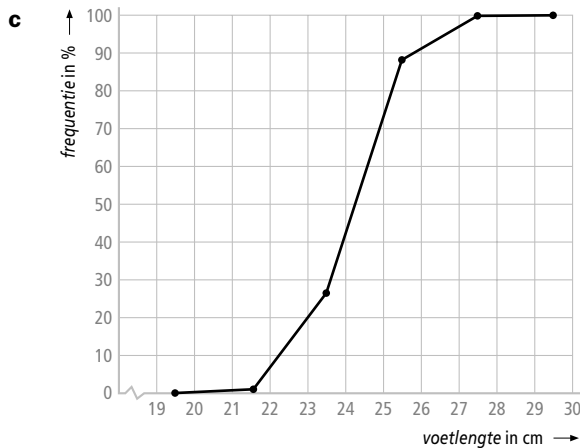


e De Dorés vertonen de grootste spreiding. Zowel de spreidingsbreedte als de kwartielafstand is groter dan bij de Bintjes.

- f De Dorés zijn de grotere aardappelloort en geven de zakken een grotere spreiding dan bij Bintjes: een aardappel meer of minder maakt bij Dorés meer verschil in gewicht dan bij Bintjes.

26a In de klasse 20-21 cm vallen alle voetlengtes van 19,5 tot 21,5 cm. De klassenbreedte is $21,5 - 19,5 = 2$ cm.

voetlengte in cm	frequentie	relatieve frequentie	relatieve somfrequentie
20-21	54	1,1%	1,1%
22-23	1282	25,6%	26,7%
24-25	3065	61,3%	88,0%
26-27	578	11,6%	99,6%
28-29	21	0,4%	100,0%
totaal	5000	100,0%	



- d Bij 50% lees je af dat de mediaan ongeveer 24,3 cm is.
- e De klasse 20-21 bevat voetlengtes van 19,5 tot 21,5 cm. Stel de klassenbreedte van 2 cm op 100% dan is 85% hiervan $0,85 \times 2 = 1,7$ cm. Tel dit op bij het linkergrens van de klasse om de voetmaten te krijgen tot 85% en je krijgt $19,5 + 1,7 = 21,2$ cm.
- f Eerst bereken je het gemiddelde op basis van de klassenmiddens en frequenties. Dit gemiddelde is $\frac{54 \times 20,5 + 1282 \times 22,5 + 3065 \times 24,5 + 578 \times 26,5 + 21 \times 28,5}{5000} \approx 24,192$ cm.

De afwijking van 3 cm betekent dat de voetlengte tussen 21,192 cm en 27,192 cm ligt. De klassen 22-23 en 24-25 vallen volledig binnen dit interval. Dit komt overeen met $1282 + 3065 = 4347$ vrouwen. Uit de klasse 20-21 moet je nog $\frac{21,5 - 21,192}{2} \times 54 \approx 8$ vrouwen bijtellen en uit de klasse 26-27 nog $\frac{27,192 - 25,5}{2} \times 578 \approx 489$ vrouwen. De grootte van de doelgroep is $4347 + 8 + 489 = 4844$ vrouwen, ongeveer 96,9% van het totaal.

- 27a** De meest voorkomende score is 5,4. Dat is dan de modus. De mediaan is het gemiddelde van de middelste twee scores (de 10^e en 11^e), dus $\frac{5,6 + 6,2}{2} = 5,9$.
- b De spreidingsbreedte is $8,2 - 3,4 = 4,8$.
- c De som van alle resultaten is 116. Het gemiddelde is $\frac{116}{20} = 5,8$.

- d** De som van alle cijfers zal nu $20 \times 0,4 = 8$ hoger komen te liggen en dus wordt het gemiddelde $\frac{116+8}{20} = 6,2$. Het gemiddelde is dus ook 0,4 punt hoger geworden. Alle scores worden met 0,4 opgehoogd, dus ook de 11^e en de 12^e. De mediaan wordt nu het gemiddelde van de nieuwe 11^e score en de nieuwe 12^e score, dus $\frac{6,0+6,6}{2} = 6,3$. Dus ook 0,4 hoger dan de oude mediaan.
- e** Nee, de laagste waarde wordt evenveel verhoogd als de hoogste waarde. Het verschil ofwel de spreidingsbreedte blijft dan gelijk.
- f** De som van de scores is $116 \times 1,1 = 127,6$ en het gemiddelde $\frac{127,6}{20} = 6,4$. De mediaan is $\frac{5,6 \times 1,1 + 6,2 \times 1,1}{20} = \frac{6,16 + 6,82}{20} = 6,49 \approx 6,5$. De spreidingsbreedte is $8,2 \times 1,1 - 3,4 \times 1,1 = 5,28 \approx 5,3$.

28a

cijfer	verschil met het gemiddelde (5,8)
3,4	2,4
3,8	2,0
4,0	1,8
4,0	1,8
4,8	1,0
5,0	0,8
5,4	0,4
5,4	0,4
5,4	0,4
5,6	0,2
6,2	0,4
6,2	0,4
6,6	0,8
6,6	0,8
6,8	1,0
6,8	1,0
7,0	1,2
7,4	1,6
7,4	1,6
8,2	2,4
totaal	22,4

- b** Het gemiddelde van de verschillen met het gemiddelde is $\frac{22,4}{20} = 1,12$.
- c** Nee. Als alle cijfers met 0,4 worden verhoogd schuift het gemiddelde ook 0,4 omhoog. Het verschil voor elk cijfer met het gemiddelde blijft gelijk en je krijgt weer 22,4 voor het totaal en 1,12 voor het gemiddelde verschil met het gemiddelde.
- d** Ja. Als alle cijfers met 1,1 worden vermenigvuldigd wordt het gemiddelde 1,1 keer zo groot. Het verschil voor elk cijfer met het gemiddelde wordt ook 1,1 keer zo groot en het gemiddelde verschil wordt dan $\frac{22,4 \times 1,1}{20} = 1,232$. Dus ook dit gemiddelde wordt dus 1,1 keer zo groot.

bladzijde 172

29a Haar gemiddelde cijfer is $\bar{X} = \frac{6,7+6,8+7,0+7,2+7,2+7,6+7,9}{6} = \frac{50,4}{6} = 7,2$.

cijfer	$X - \bar{X}$	$(X - \bar{X})^2$
6,7	-0,5	0,25
6,8	-0,4	0,16
7,0	-0,2	0,04
7,2	0,0	0,00
7,2	0,0	0,00
7,6	0,4	0,16
7,9	0,7	0,49

- c** Het gemiddelde van de kolom $X - \bar{X}$ is 0. Als je bij de verschillen met het gemiddelde het teken in acht neemt dan is het gemiddelde van deze verschillen altijd nul.
- d** Het gemiddelde van de kolom $(X - \bar{X})^2$ is $\frac{0,25+0,16+0,04+0+0+0,16+0,49}{7} \approx 0,1571$. De wortel hieruit is ongeveer 0,3964.
- e** Dit is een maat voor de spreiding rond het gemiddelde. Hoe groter het verschil tussen een waarneming en het gemiddelde hoe groter de bijdrage van die waarneming aan deze spreidingsmaat. De verschillen kunnen elkaar niet meer opheffen zoals bij onderdeel c het geval was.
- f** Door het kwadrateren worden de bijdragen aan de spreidingsmaat in de tweede kolom steeds positief.

30a -

- b** Voor het gemiddelde staat het symbool \bar{X} . Voor de standaardafwijking staat het symbool σ_x .
- c** $n = 10$ betekent dat het aantal ingevoerde waarnemingen gelijk is aan 10.
- d** Vul lijst L_1 met de cijfers van Jacolien en gebruik optie 1-Var Stats L_1 .

```
1-Var Stats
x̄=7.2
Σx=50.4
Σx²=363.98
Sx=.4281744193
σx=.3964124836
↓n=7
```

Hier zie je ook dat $\bar{X} = 7,2$ en $\sigma_x \approx 0,3964$.

bladzijde 173

31a Voer in lijst L_1 in de voorkomende jaarsalarissen en gebruik lijst L_2 voor bijbehorende frequenties. Met optie 1-Var Stat L_1, L_2 krijg je:

```
1-Var Stats
x̄=52666.66667
Σx=790000
Σx²=6.85E10
Sx=43828.67076
σx=42342.51869
↓n=15
```

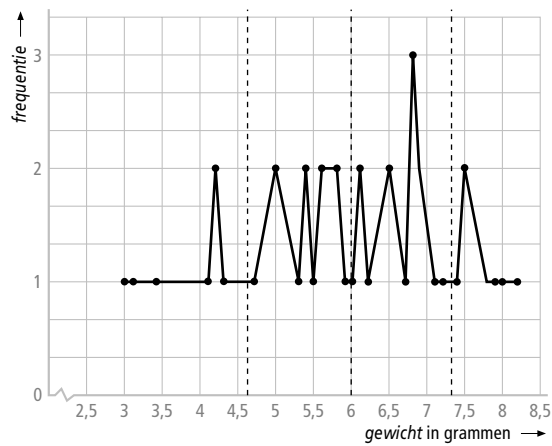
Je leest af $\bar{X} \approx 52666,67$ en $\sigma_x \approx 42342,52$

b,c,d,e	verandering	gevolg voor het gemiddelde	gevolg voor de standaardafwijking
	frequenties $\times 10$	blijft zoals het was	blijft zoals het was
	salarissen $\times 5$	wordt $5\times$ zo groot	wordt $5\times$ zo groot
	frequenties $+ 10$	wordt ongeveer 65273	wordt ongeveer 48274
	salarissen $+ 10000$	wordt 10000 meer	blijft zoals het was

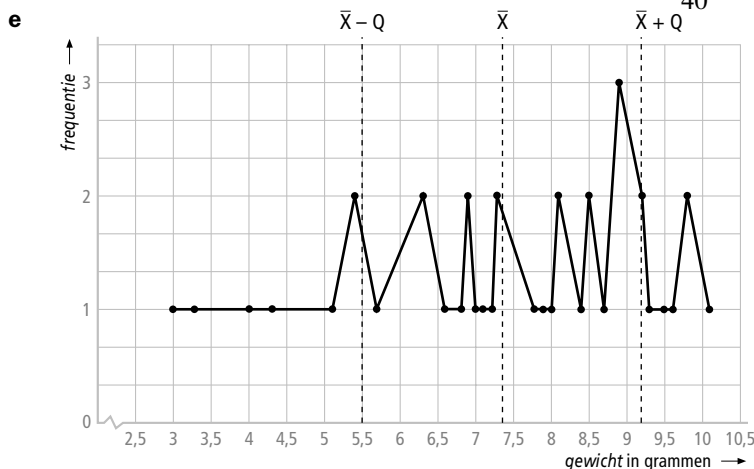
32a Acht van de 25 pakken zijn lichter dan 100 gram, dit is $\frac{8}{25} \times 100\% = 32\%$.

- b** De pakjes met 105 en 106 gram komen het meest voor. De pakjes met 105 gram liggen bovendien het meest centraal van de verdeling, dus heeft de fabrikant het gewicht op 105 gram ingesteld.
- c** De pakjes van 97, 98 en 99 gram zijn te licht. Dit zijn er $1+1+2 = 4$ stuks. Van de 58 pakjes is dit ongeveer 6,9%.
- d** $\bar{X} \approx 104,97$ en $\sigma_x \approx 3,32$
- e** De controledienst constateert dat de vulmachine nu beter staat ingesteld.

33a Uit het steelbladdiagram lees je de volgende frequenties van boom A af.



- b** Het gemiddelde is 5,955 gram. De standaardafwijking is ongeveer 1,342.
- c** Het gemiddelde plus de standaardafwijking is $5,955 + 1,342 = 7,297$ gram. Het gemiddelde minus de standaardafwijking is $5,955 - 1,342 = 4,613$ gram.
- d** Tussen 4,613 en 7,297 gram liggen 26 kastanjes. Dat is $\frac{26}{40} \times 100\% = 65\%$.

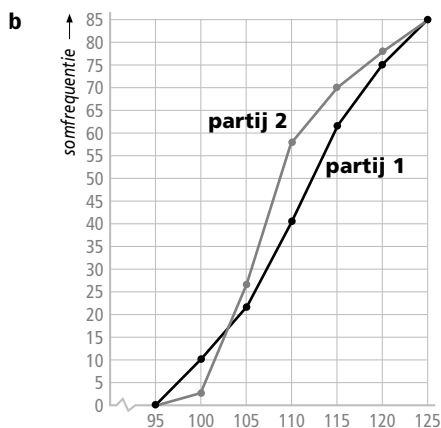


Het gemiddelde is ongeveer 7,33 gram en de standaardafwijking is ongeveer 1,866. De gemiddelde plus de standaardafwijking is $7,33 + 1,866 = 9,196$ gram. Het gemiddelde minus de standaardafwijking is $7,33 - 1,866 = 5,464$. Tussen 5,464 en 9,196 gram liggen 24 kastanjes. Dat is $\frac{24}{40} \times 100\% = 60\%$.

- f Aangenomen dat de kastanjabomen vergelijkbare grootte en ouderdom hebben kun je verwachten dat boom, met gemiddeld zwaardere kastanjes, in het park staat omdat de groeimogelijkheden voor de boom hier gunstiger zijn bij de parkeerplaats. Boom B staat dus in het park.

bladzijde 174

gewicht in grammen	partij 1	cumulatieve frequentie partij 1	partij 2	cumulatieve frequentie partij 2
[95; 100)	10	10	3	3
[100; 105)	12	22	24	27
[105; 110)	19	41	31	58
[110; 115)	21	62	12	70
[115; 120)	13	75	8	78
[120; 125)	10	85	7	85

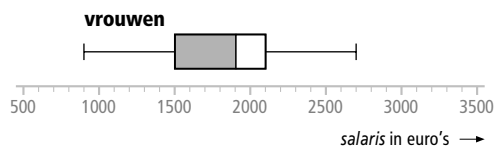


- c De mediaan ligt halverwege 0 – 85, dus bij 42,5. Voor partij 2 ligt de mediaan bij een kleiner gewicht dan voor partij 1. Partij 2 bestaat dus uit kleinere kiwi's. De somfrequentiepolygoon van partij 2 loopt rond de mediaan steiler dan voor partij 1. Partij 2 heeft dus een kleinere spreiding dan partij 1 en de kiwi's van partij 2 lijken (in ieder geval qua gewicht) meer op elkaar dan die van partij 1.

- 35a De aanduiding 'absoluut' bij de linker as geeft aan dat de as hoort bij de absolute grafiek met de rode lijn. De aanduiding 'relatief' bij de rechter as geeft aan dat de as hoort bij de relatieve grafiek met de zwarte lijn. Op de absolute as lees je de aantallen personen af die binnenlands migreerden en op de relatieve as de aantallen per 1000 personen (1000 inwoners).

- b Het absolute aantal binnenlandse migranten bedroeg $400 \times 1000 = 400\,000$ personen. Het relatieve aantal is 40 per 1000 personen.
- c De bevolking groeide tussen 1960 en 1963. Stel er migreerden 500 000 personen op de 11 miljoen personen in 1960, dan was dat relatief $\frac{500\,000}{11\,000\,000} \approx 45$ per duizend. Als het aantal migranten in 1963 hetzelfde was, maar nu op een bevolking van 12 miljoen, dan was dat relatief $\frac{500\,000}{12\,000\,000} \approx 42$ per duizend. Relatief nam het aantal dus af en daardoor daalt de grafiek.
- d In 1973 waren er 750 000 migraties en in 1963 was dat aantal 500 000. Dat is een toename van 250 000. Gemiddeld per jaar is dat een toename van 25 000 personen.

- 36a** De modale klasse is de klasse met de hoogste frequentie. Dat is hier $[1600; 1800)$.
- b In salarisgroep 9 t/m 12 vallen $5 + 3 + 0 + 0 = 8$ vrouwen van de 75. Dat is ongeveer 10,7%.
 - c Bij de mannen is dat $5 + 3 + 1 + 1 = 10$ van de 50, dat is dus 20%.
 - d Het laagste maandsalaris bij de vrouwen is het klassenmidden van de klasse $[800; 1000)$, dus €900. Het hoogste maandsalaris is het klassenmidden van de klasse $[2600; 2800)$, dus €2700. De mediaan is het salaris van de 38^e vrouw. Je vindt door te tellen dat dit is in de klasse $[1800; 2000)$. Omdat er met klassenmiddens wordt gewerkt (interpolatie wordt achterwege gelaten in deze opgave) is de mediaan dus 1900. Q_1 is het salaris van de 19^e vrouw. Zij valt in salarisklasse $[1400; 1600)$ en dus $Q_1 = 1500$. Het derde kwartiel is het salaris van de 57^e vrouw. Zij valt in salarisklasse $[2000; 2200)$ en dus is $Q_3 = 2100$.



- e Omdat $Q_3 = 2100$ verdient 25% van de vrouwen meer dan € 2100. Dat zijn ongeveer 19 vrouwen. Bij de mannen is de mediaan gelijk aan 2100, dus verdient 50% van de mannen meer dan € 2100. Dat zijn 25 mannen. De bewering is juist als je kijkt naar de relatieve aantallen ($50\% = 2 \times 25\%$). De bewering is onjuist als je kijkt naar de absolute aantallen ($25 \neq 2 \times 19$).
- f Voer de klassenmiddens in je rekenmachine in via lijst L_1 en bijbehorende aantallen mannen in lijst L_2 . Met optie 1-Var Stat L_1, L_2 lees je af dat $\bar{X} \approx 2064$ en $\sigma_x \approx 391$. Afgerond op honderdtallen krijg je $\bar{X} \approx 2100$ (euro) en $\sigma_x \approx 400$ (euro).
- g De lagere salarissen komen bij de vrouwen het meeste voor.
- h Voer in je rekenmachine als lijst L_3 in de aantallen vrouwen per salarisgroep en definieer lijst L_4 met $L_4 = L_2 + L_3$. In lijst L_4 staan dan de aantallen werknemers per salarisgroep. Met 1-Var Stat L_1, L_4 lees je af: $\bar{X} \approx 1918$ en $\sigma_x \approx 450$. Afronden op honderdtallen geeft $\bar{X} \approx 1900$ en $\sigma_x \approx 500$.

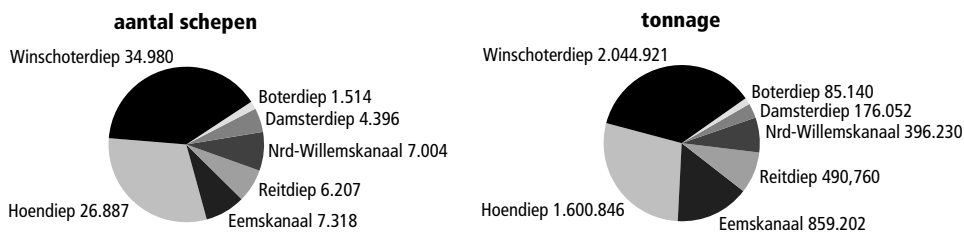
bladzijde 175

- 37a** Volgens de tabel deden 60,2% van de jongens Economie en 47,9% van de meisjes. Er deden $344 \times 0,602 = 207$ jongens en $493 \times 0,479 = 236$ meisjes. Er waren dus meer meisjes dan jongens die Economie deden.

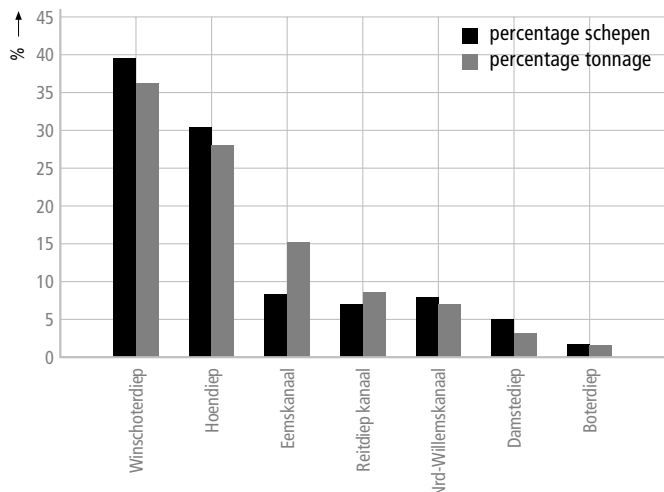
- b** Tel alle percentages bij de meisjes bij elkaar op en je vindt 519,2% aan percentages dat uitgedeeld is aan de vakken. Iedereen kiest minstens 5 andere vakken naast Nederlands, dus 500% wordt daaraan uitgedeeld. De 19,2% die overblijft kan alleen maar aan het zesde vak worden uitgedeeld. Het gaat daarbij dan om $493 \times 0,192 = 95$ meisjes.
- c** De rechter figuur lees je bij Economie af dat 7,5% van de jongens het vak niet weer zou kiezen. In opdracht a vond je dat 207 jongens Economie deden, dus $207 \times 0,075 = 16$ hiervan zou het vak niet weer kiezen. In de linker figuur lees je af dat van de jongens die een ander vakkenpakket zouden kiezen er 34% de voorkeur aan Economie zou geven. Er zijn 127 jongens die een ander vakkenpakket zouden kiezen, dat zijn er $127 \times 0,34 = 43$. Als de jongens achteraf het gewenste pakket hadden gekozen zouden er dus $207 - 16 + 43 = 234$ jongens voor Economie kiezen. Bij de meisjes voer je dezelfde soort berekening uit maar lees je de percentages 17,5% en 23% af. Je krijgt dat $236 - 236 \times 0,175 + 232 \times 0,23 = 236 - 41 + 53 = 248$ meisjes die achteraf voor Economie zouden kiezen. Er zijn dus nog steeds meer meisjes dan jongens die Economie zouden doen.

bladzijde 176

- I-1a** Klik op het cirkeldiagram-icoon in de werkbalk. Dubbelklik op ‘aantal schepen’ en klik op OK. Met vormgeving kun je bijvoorbeeld het tonen van labels regelen.



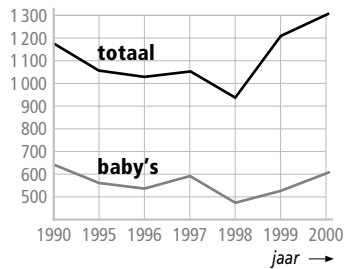
- b** Klik op het staafdiagram-icoon in de werkbalk. Dubbelklik op ‘percentage schepen’ en ‘percentage tonnage’ en klik op OK. Onder de knop vormgeven kun je een aantal instellingen wijzigen voor de weergave. Vergroot eventueel het venster om alle gegevens duidelijk in beeld te brengen.



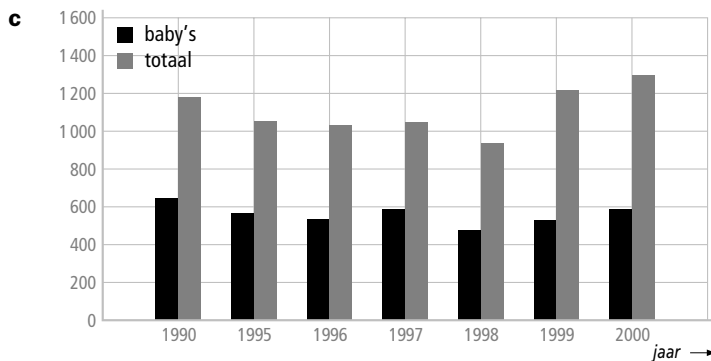
Het Hoendiep heeft een aandeel van 30,45% bij het aantal schepen en 28,32% bij het tonnage. Het aandeel bij het aantal schepen is dus groter.

- c Het percentage bij het tonnage is voor het Eemskanaal ongeveer twee maal zo groot als het percentage bij het aantal schepen. De andere kanalen hebben ongeveer gelijke percentages hiervoor. Door het Eemskanaal voeren dus schepen die groter waren en/of zwaarder beladen waren dan de schepen die in de andere kanalen voeren.

I-2a Klik op het lijndiagram -icoon in de werkbalk. Dubbelklik op 'baby's' en 'totaal' en klik op OK.



b De data worden per jaar gegeven en niet per maand.

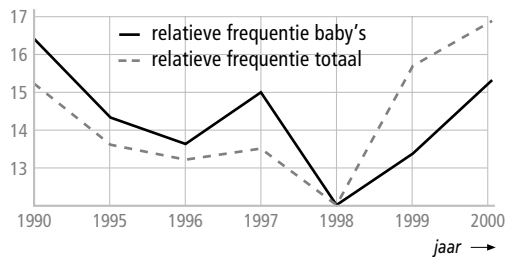


d Het lijndiagram leest beter af voor de ontwikkeling in de tijd.

e Er wordt automatisch een kolom gecreëerd met relatieve frequenties.

f Het getal 16.33 bij 1990 betekent dat in 1990 het percentage geadopteerde baby's 16,33% was van het totaal aantal geadopteerde baby's gedurende 1990 – 2000.

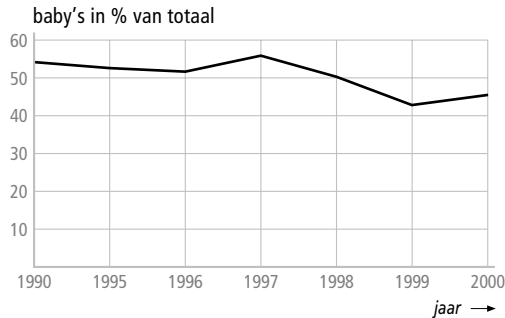
g Dat het verloop grillig lijkt, komt met name doordat de verticale as niet bij nul begint en wordt het verschil van een paar procent dat er is tussen jaren uitgerekt over de hele figuur.



bladzijde 177

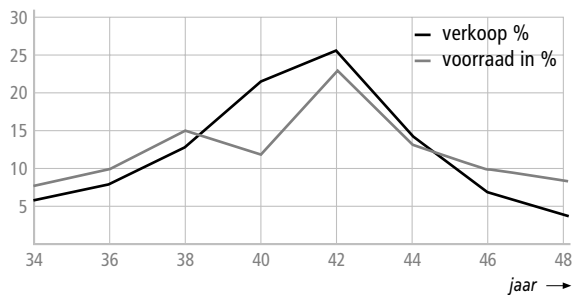
I-3a,b Je vergelijkt de jaarlijkse babyadopties met het totaal aantal baby's door de verhouding te berekenen. Dat doe je door het getal uit de kolom 'baby's' te delen door het getal uit de kolom 'totaal' bij elk jaar. Om de verhouding in procenten weer te geven vermenigvuldig je nog met 100.

- c Wijzig het minimum en het maximum voor de linkeras in 0 en 60 bij vormgeven > Assen > As links > Verander en vink Automatisch uit. De grafiek ziet er dan niet meer zo grillig uit.



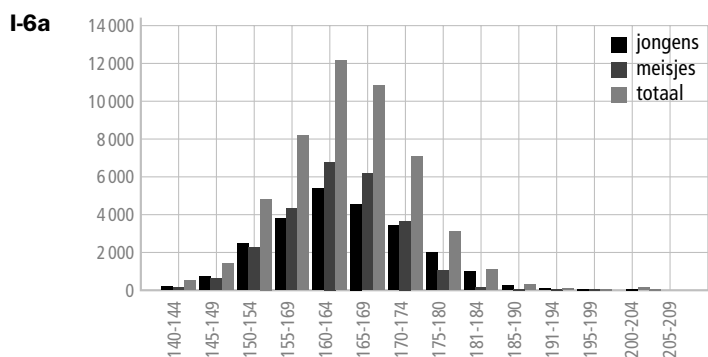
- d Het beleid zal wellicht zijn dat het bij 50% van de adopties om baby's gaat.

- I-4a Klik op het lijndiagram-icoon in de werkbalk. Kies 'maat' voor de horizontale as, dubbelklik op 'verkoop %' en 'voorraad %' en klik op OK. Commentaar: De voorraad voor maat 40 is niet te groot, terwijl de voorraad bij de grotere maten wat kleiner kan.



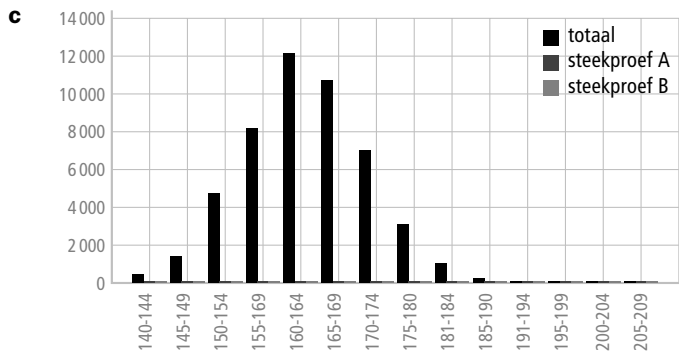
- b Verhoog het aantal damespantalons met maat 40 en 42.

- I-5a Het lijkt alsof de spoorwegen steeds veiliger werden. De grafiek begint niet bij nul, daarom lijkt het alsof er vanaf 1960 bijna helemaal geen ongelukken meer waren.
 b Het verloop van de aantallen treinreizigers. Het aantal ongelukken tussen 1920 en 1960 werd bijna 3 keer kleiner, maar het aantal treinreizigers nam in de VS nog veel sterker af. Relatief gezien werden de spoorwegen dus juist onveiliger.



Binnen elk groepje geldt: staaf jongens + staaf meisjes = staaf voor het totaal.

- b In de klasse 160 – 164 vallen de meeste jongens en meisjes. In klas 1 en 2 zijn de leerlingen rond de 13 jaar. De jongens blijven op die leeftijd over het algemeen wat achter bij de groei van de meisjes.



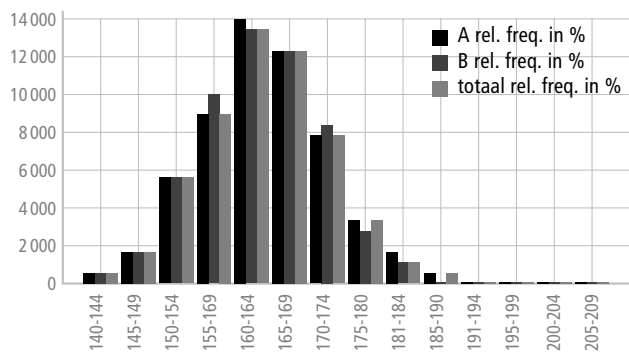
De staven behorende bij een steekproef zijn erg klein ten opzichte van het totaal en daarom niet afleesbaar.

- d** Klik op de kop van een lege kolom voor het opgeven van een nieuwe variabele. Creëer de variabelen 'A rel freq in %', 'B rel freq in %' van het type Geheel getal. Klik weer op de kop van de kolom en kies 'Omrekenen...'. Voer de volgende formules in:

$$A \text{ rel freq in \%} = \left[\text{steekproef A} \right] / 500 * 100$$

$$B \text{ rel freq in \%} = \left[\text{steekproef B} \right] / 500 * 100$$

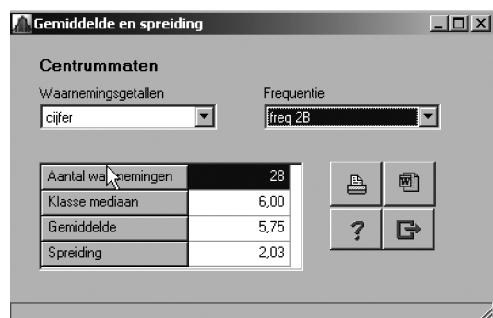
$$\text{totaal rel freq in \%} = \left[\text{totaal} \right] / 50000 * 100$$



Beide steekproeven geven een goed beeld. Alle kenmerken van het totaal zijn in beide steekproeven goed te zien: de vorm van de verdeling en de verhouding tussen de klassen.

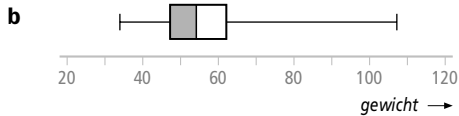
bladzijde 178

- I-7a** Kies opties > Centrummaten en kies als waarnemingsgetallen 'cijfer' en als 'Frequentie' freq 2B.



- b Het blijkt dat de mediaan en het gemiddelde identiek zijn aan die van klas 2A. Het verschil zit hem in de spreiding rond het gemiddelde. De spreiding van klas 2B is behoorlijk wat groter dan die van klas 2A.

I-8a Van 92 kinderen is een aantal gegevens verzameld: naam, lichaamslengte, gewicht, geslacht, brildracht, vrijetijdsbesteding.



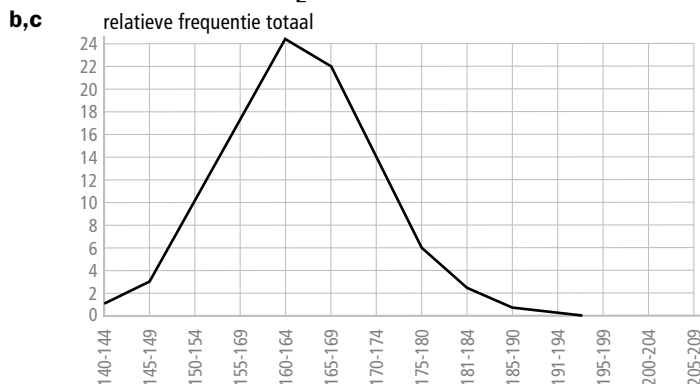
Wat opvalt dat met name de spreidingsbreedte erg groot is. Dat zou natuurlijk kunnen worden veroorzaakt door wat uitschieters naar boven.

- c Na sortering blijkt dat van 4 kinderen geen gewicht bekend is. Er zijn dus slechts 88 waarnemingen. Het laagste gewicht blijkt 34 kg, Q_1 is bij 88 kinderen de mediaan van de eerste 44 waarnemingen. Je moet dus het gemiddelde tussen de 22^e en 23^e waarneming bepalen: $Q_1 = \frac{47+48}{2} = 47,5$. De mediaan is het gemiddelde van de 44^e en 45^e waarneming, dat is 54,5 kg. Het derde kwartiel Q_3 is hier het gemiddelde van de 66^e en 67^e waarneming, dan is hier 62,5 kg. Het hoogste gewicht dat voorkomt is van Christian. Hij is 108 kg.
- d Het verschil tussen het hoogste en het laagste gewicht is $108 - 34 = 74$ kg en het verschil tussen Q_1 en Q_3 is $62,5 - 47,5 = 15$ kg.

I-9 De spreidingsbreedte en de kwartielafstand is bij I-8 bepaald, ook al zijn de termen niet gebruikt. De spreidingsbreedte is 74 kg en de kwartielafstand 15 kg.

bladzijde 179

I-10a De eerste klasse die in de data voorkomt is 140 – 144 . Deze klasse bevat alle lengtes die na afronding in cm 140, 141, 142, 143 of 144 opleveren. De eigenlijke klasse van lichaamslengtes is dan dus $[139,5; 144,5)$ en de klassenbreedte is $144,5 - 139,5 = 5$ cm.



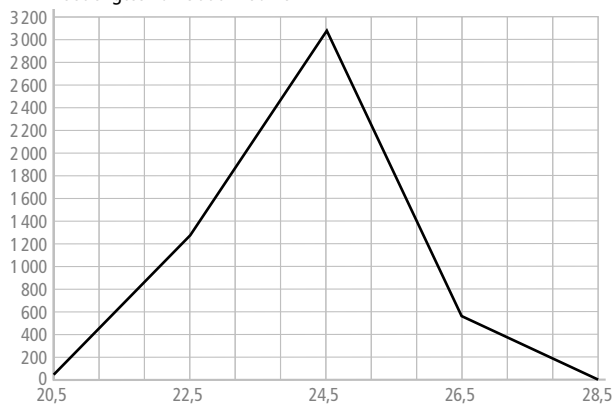
- d De verdeling is redelijk symmetrisch. Een schatting zou kunnen zijn: iets (bijvoorbeeld 40% van een klassenbreedte, dus 2 cm) naar rechts ten opzichte van de klasse 160 – 164. Je moet hier natuurlijk rekenen met klassenmiddens. De schatting wordt dan $162 + 2 = 164$ cm.

- e In de klasse 180 – 184 vallen in totaal 1155 kinderen. Alle kinderen die minstens 179,5 cm zijn en korter dan 184,5 vallen in deze klasse. 10% van de totale klasse behoort bij lichaamslengtes kleiner dan 180. Als je ervan uitgaat dat de kinderen gelijkmatig over de klasse zijn verdeeld mag je dus verwachten dat 10% van het totaal aantal kinderen (1155) een lengte hebben van minder dan 180 cm. En 10% van 1155 is ongeveer 115.
- f De klassen 165 – 169, 170 – 174, 175 – 179, 180 – 184 en 184 – 189 behoren daar steeds in hun geheel bij. Van de klasse 160 – 164 moet je 90% meenemen en van de klasse 190 – 194 voor 10%. Het geschatte aantal kinderen met een lengte tussen 160 en 190 cm is $10869 + 7117 + 3142 + 1155 + 335 + \frac{9}{10} \times 12152 + \frac{1}{10} \times 103 \approx 33565$. Dit is ongeveer 67,1%.

I-11a Bijvoorbeeld de klasse 20 – 21 omvat alle voetlengtes van het interval $[19,5; 21,5)$. De klassenbreedte is 2 cm.

- b Voer de data in en kies geschikte namen voor de variabelen, bijvoorbeeld ‘klassenmidden’ en ‘frequentie’.

c voetlengtes van 5000 vrouwen



- d De totale klasse bestaat uit voetlengtes uit het interval $[19,5; 21,5)$. Vrouwen in dit interval die een voetlengte hebben minder dan 21,2 cm vallen dus in het interval $[19,5; 21,2)$ en dat is $\frac{1,7}{2} \times 100\% = 85\%$ van de totale klasse. Als de vrouwen gelijkmatig over een klasse zijn verdeeld, dan heeft dus ook 85% van de vrouwen in die klasse een voetlengte kleiner dan 21,2 cm.
- e Gebruik optie centrummaten, neem als waarnemingsgetallen de klassenmiddens en klik de frequentievariabele aan. Je ziet dan dat het gemiddelde gelijk is aan 24,192, dus ongeveer 24,2 cm. Een vrouw die qua voetlengte hoogstens 3 cm van het gemiddelde 24,2 afwijkt heeft een voetlengte tussen 21,2 en 27,2 cm. De vrouwen uit de klassen 22 – 23 en 24 – 25 voldoen daaraan, uit de klasse 20 – 21 geldt dat voor ongeveer 15% en uit de klasse 26-27 geldt dat ongeveer voor 85%. Het gaat dan bij benadering om $0,15 \times 54 + 1282 + 0,85 \times 578 \approx 4846$ vrouwen, dus ongeveer 96,9%.

I-12a Met optie Meer Statistiek > Centrummaten lees je af dat het gemiddelde 5,8 is, de mediaan 5,9 en de modus 5,4.

- b De spreidingsbreedte is de hoogste minus de laagste score = maximum – minimum = $8,2 - 3,4 = 4,8$.
- c Zowel het gemiddelde als de mediaan zullen 0,4 hoger komen te liggen.
- d -
- e Omdat alle cijfers, dus ook het maximum en minimum met 0,4 worden verhoogd, blijft de spreidingsbreedte = maximum – minimum ongewijzigd.
- f Alle centrummaten, dus gemiddelde, mediaan en modus worden met een factor 1,1 vermenigvuldigd.

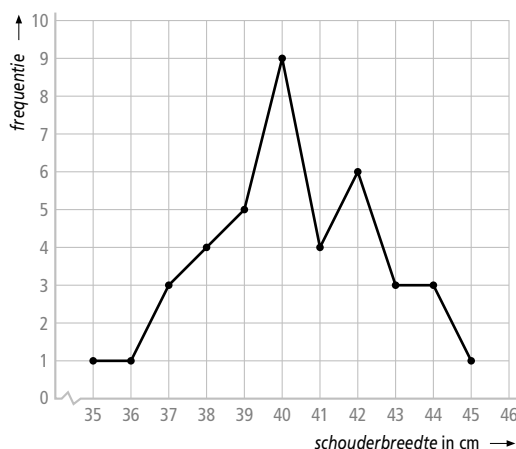
- I-13a** Met optie Meer statistiek > Centrummaten en de keuze voor de variabele gem-cijfer vind je uiteraard een gemiddelde van precies 0.
- b** De derde kolom is identiek aan de tweede, alleen het teken is weggelaten.
 - c** VU-Stat geeft als gemiddelde van de derde kolom 1,12. Het gaat hier om het gemiddelde absolute verschil met het gemiddelde.
 - d** Deze zal niet veranderen. De verschillen tussen de cijfers en gemiddelde blijven ongewijzigd.
 - e** Omdat alle verschillen met gemiddelde met een factor 1,1 worden vermenigvuldigd geldt dat ook voor gemiddelde verschil met het gemiddelde.

bladzijde 182

- T-1a** Ja, de gegevens in de grafiek zijn relatief. Ze zijn weergegeven per 1000 inwoners of per 100 levendgeborenen dus hebben ze betrekking op een ander aantal en zijn daarom relatief. Je kunt de absolute aantallen niet aflezen.
- b** In de grafiek heeft de sterftelijn een piek bij de Spaanse griep. De piek ligt ongeveer 5 boven het normale niveau als je de sterftelijn gewoon zou doortrekken. De sterfte door de Spaanse griep was dus 5 per 1000 inwoners. Bij 6 625 300 inwoners waren dat $\frac{6\,625\,300}{1000} \times 5 = 33\,130$ mensen.
 - c** Uit de bovenste lijn lees je af dat er 25 levendgeborenen per 1000 inwoners waren. Dat zijn bij 6625300 inwoners dus $\frac{6625300}{1000} \times 25 = 165\,633$ levendgeborenen. Volgens de onderste lijn overleden 10 levendgeborenen per 100, dus totaal $\frac{165\,633}{100} \times 10 = 16\,563$ levendgeborenen.

- T-2a** De kleinste schouderbreedte was 35,2 en de grootse 45,3 cm. Dat is een verschil van bijna 10 cm. De elf klassen krijg je door een indeling per centimeter te maken.

schouderbreedte in cm	frequentie
[34,5; 35,5)	1
[35,5; 36,5)	1
[36,5; 37,5)	3
[37,5; 38,5)	4
[38,5; 39,5)	5
[39,5; 40,5)	9
[40,5; 41,5)	4
[41,5; 42,5)	6
[42,5; 43,5)	3
[43,5; 44,5)	3
[44,5; 45,5)	1

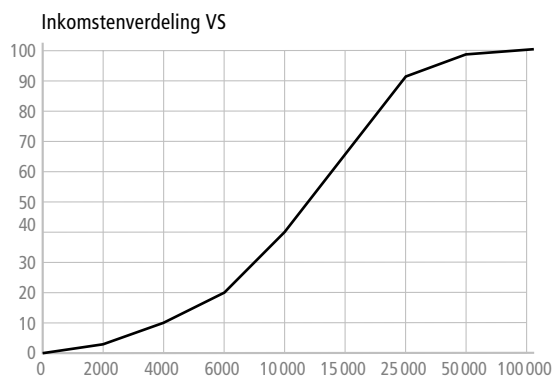


- b De verandering van 9 naar 4 tussen de opeenvolgende klassen $[39,5; 40,5)$ en $[40,5; 41,5)$ is met een verschil van 5 het grootst.
- c De schatting moet je berekenen. Het blijft echter een schatting want je maakt gebruik van de klassenmiddens als benadering in plaats van de waarnemingsgetallen. Het gemiddelde van de 40 schouderbreedtes is bij benadering
- $$\frac{35 \times 1 + 36 \times 1 + 37 \times 3 + 38 \times 4 + 39 \times 5 + 40 \times 9 + 41 \times 4 + 42 \times 6 + 43 \times 3 + 44 \times 3 + 45 \times 1}{40} = \frac{1611}{40} \approx 40,3 \text{ cm.}$$

T-3a De frequentietabel met cumulatieve relatieve frequenties wordt:

<i>inkomen</i> in dollars	<i>bevolkings-</i> <i>percentage</i>	<i>cumulatieve</i> <i>bevolkingspercentage</i>
$[0; 2000)$	3	3
$[2000; 4000)$	7	10
$[4000; 6000)$	10	20
$[6000; 10000)$	20	40
$[10000; 15000)$	26	66
$[15000; 25000)$	26	92
$[25000; 50000)$	7	99
$[50000; 100000)$	1	100

- b Bij een somfrequentiepolygoon worden de cumulatieve frequenties boven de rechter klassengrenzen gezet.



- c Uit de somfrequentiepolygoon kun je aflezen dat naar schatting 80% van de bevolking minder dan 20 000 dollar verdient. Deze schatting is gebaseerd op de veronderstelling dat de salarissen gelijkmatig over de klasse $[15000; 25000)$ zijn verdeeld.

bladzijde 183

- T-4a Dat is ongeveer vanaf dag 3 tot dag 22, dus 20 dagen.
- b In die periode zal de werkbij ongeveer 10% van haar tijd aan besteden aan 'raat sluiten'. Dit komt overeen met de gemiddelde lengte van de staafjes bij 'raat sluiten' in de actieve periode.
- c Door alle staafjes te schatten en op te tellen en vervolgens de som te delen door 24 kom je ongeveer op 9%.

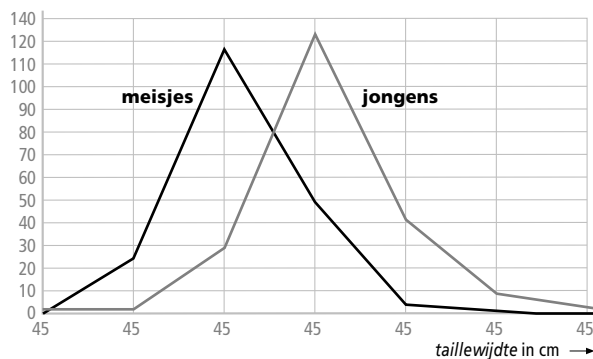
- d 'Cellen poetsen' lijkt de omvangrijkste taak, al zal de hoeveel tijd te besteden aan van 'raat sluiten' in de buurt komen. Het gaat er natuurlijk om de taak te vinden waarbij de oppervlakte van het staafdiagram het grootst is.

- T-5a,b** Voor de drie series geldt dat de spreidingsbreedte steeds $100 - 0 = 100$ is. Toch is er verschil in mate van spreiding: De grootste spreiding heeft serie B, omdat de waarnemingsgetallen voor het overgrote deel binnen het interval $[0; 100]$ extreem ver van het gemiddelde 50 liggen. Daarna komt serie A, met waarnemingsgetallen die gelijkmatig over het interval zijn verdeeld. Tenslotte heeft serie C van de drie series de minste spreiding. Afgezien van twee uitschieters, namelijk 0 en 100, ligt alles dicht op elkaar rond het gemiddelde 50.
- c Bij 7 waarnemingsgetallen is Q_1 gelijk aan de 2^e waarneming en is Q_3 gelijk aan de 6^e waarneming. Voor de standaardafwijkingen is het wellicht gemakkelijk om de series in lijsten L_1 , L_2 en L_3 in te voeren en gebruik te maken van optie 1-Var Stat.

	kwartielafstand	standaardafwijking
A	60	31,6
B	98	45,4
C	4	26,9

Zowel de kwartielafstand als de standaardafwijking sluiten hier aan bij het intuïtieve begrip dat we bij onderdeel b hebben gevolgd. In ieder geval zetten deze spreidingsmaten de series qua mate van spreiding in dezelfde volgorde als bij opdracht b.

T-6a



- b De frequentiepolygonen hebben dezelfde vorm. Die van de jongens lijkt naar rechts verschoven ten opzichte van die van de meisjes.
- c Voer de klassenmiddens en de frequenties in de lijsten L_1 , L_2 en L_3 zoals hieronder.

L1	L2	L3	3
45	0	2	
55	25	2	
65	116	29	
75	49	123	
85	4	41	
95	1	9	
105	0	3	

$L3(1)=2$

Met 1-Var Stat L_1 , L_2 lees je af $Q_1 = 65$, Mediaan = 65, $Q_3 = 75$ voor de meisjes. Dus de kwartielafstand is 10. Met 1-Var Stat L_1 , L_3 lees je af $Q_1 = 75$, Mediaan = 75, $Q_3 = 85$ voor de jongens. Ook hier is de kwartielafstand 10. Een betere methode veronderstelt per klasse een gelijkmatige verdeling van de taillewidthtes.

Dit gaat als volgt. Bij de meisjes is Q_1 gelijk aan de mediaan van de eerste 97 waarnemingen. Je moet dus de 49^e waarneming hebben. De 49^e waarneming is de 24^e waarneming in de klasse $[60; 70)$. Bij gelijkmatige verdeling van de 116 waarnemingen in die klasse, is $Q_1 \approx 60 + \frac{24}{116} \times 10 \approx 62,1$. Met interpolatie kun je de mediaan en het derde kwartiel bepalen.

	Q_1	mediaan	Q_3	kwartielafstand
meisjes	62,1	66,3	71,0	8,9
jongens	71,5	75,9	80,0	8,5

- d** Met 1-Var Stat L_1, L_2 lees je af $\bar{X} \approx 66,8$ en $\sigma_x \approx 6,9$ voor de meisjes. Met 1-Var Stat L_1, L_3 lees je af $\bar{X} \approx 76,4$ en $\sigma_x \approx 8,6$ voor de jongens.
- e** Onze schattingen liggen voor een groot deel in dezelfde orde van grootte als de gemiddelden en standaardafwijkingen behorende bij de oorspronkelijke data. Dat geldt ook ten aanzien van de geschatte medianen. Alleen de geschatte kwartielafstanden wijken nogal af.
- f** Nee, dat kun je niet zomaar concluderen. Je moet de taille zien in relatie tot de lichaamslengte en ook de vorm van het lichaam. Kortom, de taillewijdte is niet een goede maat dikte of vetzucht.
- T-7a** Klas A zal vermoedelijk wat meer onvoldoendes hebben dan klas B. Dit omdat de cijfers van klas A een grotere spreiding hebben om het gemiddelde 6,2 dan klas B en omdat 6,2 toch niet heel ver van de voldoende grens afluigt.
- b** De frequentiepolygoon van klas B zal spitsers zijn en de somfrequentiepolygoon van klas B zal steiler zijn.