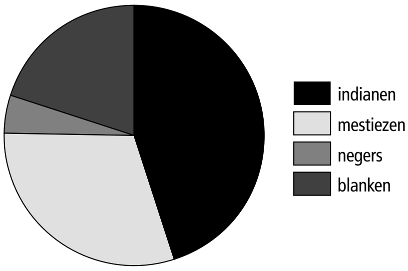
Hoofdstuk 7 - Statistische verwerking

bladzijde 151

V-1 De totale bevolking in Latijns-Amerika omvatte rond 1880 19,6 miljoen mensen.
Hiervan behoorden $\frac{7,6}{16,9} \cdot 100\% \approx 45\%$ tot de indianen.
De bijbehorende sectorhoek is dan $0,45 \cdot 360° \approx 162°$.
Op soortgelijke manier vind je de andere sectorhoeken:
mestiezen: $\frac{5,1}{16,9} \cdot 360° \approx 109°$, negers: $\frac{0,8}{16,9} \cdot 360° \approx 17°$ en blanken: $\frac{3,4}{16,9} \cdot 360° \approx 72°$.



V-2a De sterkste daling van het jaargemiddelde vond plaats in de periode 1993 – 1994.
In het jaar 1994 daalde de notering dus het sterkst.

b De sector BTW is goed getekend want 6% van 360° is $0,06 \cdot 360° = 21,6°$ en de sectorhoek is ongeveer 22°.

c De sectorhoek bij de Leeuwardernotering is ongeveer 178° en dat komt overeen met $\frac{178}{360} \cdot 100\% \approx 49,4\%$ van de winkelprijs.

d $49,4\%$ van $f\ 11,96$ is $11,96 \times 0,494 \approx 5,91$.
Dit komt overeen met de prijs van de Leeuwardernotering voor het jaar 1995, die uit de linkergrafiek is af te lezen: ongeveer $f\ 5,90$.

V-3a Van merk B zijn de meeste blikjes onderzocht: 25 van merk A en 34 van merk B.

b Merk A: 54, 60, 61, 61 en 61 gram zijn lichter, dat is $\frac{5}{25} \cdot 100\% = 20\%$.
Merk B: 58, 62, 62 en 64 gram zijn lichter, dat is $\frac{4}{34} \cdot 100\% \approx 11,8\%$.

c Merk A: gemiddeld $\dfrac{54+60+61+61+..........+91+91+91}{25} = \dfrac{1865}{25} = 74,6$ gram.

Merk B: gemiddeld $\dfrac{58+62+62+64+..........+90+91+98}{34} = \dfrac{2535}{34} \approx 74,6$ gram.

bladzijde 152

1a Van deze groep van 570 personen gaat 45% nooit naar een concert.
Daaruit kun je niet concluderen, dat dit percentage ook van toepassing is op de totale Rotterdamse bevolking.

b Ja, het is mogelijk dat in Rotterdam 80% van de mensen nooit naar een concert gaat.
Als de groep van 570 personen een goede afspiegeling is van de Rotterdamse bevolking dan is zo'n uitslag echter niet erg waarschijnlijk.

c Nee, want de muzikale belangstelling kan in Utrecht anders zijn dan in Rotterdam.

2a De steekproef is getrokken uit alle leden van de omroepvereniging.
b De grootte van de steekproef is 2000 want er zijn 2000 formulieren verstuurd.
c Van de teruggestuurde antwoorden was wel de meerderheid positief, maar er is niets bekend over de mening van de leden die niet geantwoord hebben.
Bovendien is maar een deel van de vereniging in de steekproef opgenomen.
d Nee. Het is te verwachten, dat de leden van een bepaalde omroepvereniging positiever zijn over de eigen programma's dan de Nederlanders in het algemeen.
e Nee, want er kunnen verschillende redenen zijn geweest waarom de leden het formulier niet hebben teruggestuurd.
f Er zijn $1143 \times 25 = 28\,575$ data verzameld.

bladzijde 153

3 Op de beschreven manier heeft elk lid van de omroepvereniging evenveel kans om in de steekproef opgenomen te worden.

4a De enquête is niet aselect want mensen met een geheim telefoonnummer of mensen die niet thuis zijn op het moment van het onderzoek vallen buiten de steekproef.
Bovendien is het onderzoek niet anoniem waardoor mensen wellicht terughoudend zullen antwoorden.
b Mensen die vroeg naar hun werk gaan vallen buiten de steekproef.
Ook andere categorieën zoals bejaarden in verzorgingshuizen bijvoorbeeld blijven buiten beeld.
c In de meeste gezinnen is het dan spitsuur (maaltijd, kinderen naar bed etc.) waardoor er een geringere bereidheid is om deel te nemen en de antwoorden wellicht snel en minder doordacht worden gegeven.
Antwoorden worden beïnvloed omdat anonimiteit is niet gewaarborgd.
d Alcoholisten zullen misschien vaker dan anderen geneigd zijn om het formulier niet terug te sturen en de categorie analfabeten valt er ook buiten.

5a $3983 + 1760 = 5743$.
b $\frac{1760}{5743} \cdot 100\% \approx 30{,}6\%$.
c $\frac{1806}{(2849 + 1806)} \cdot 100\% \approx 38{,}8\%$.
d Ja, want het percentage niet-beantwoorders is toegenomen van 30,6% tot 38,8%, dus het percentage meewerkenden is afgenomen van 69,4% tot 61,2%.
e In 2003 weigerde 69% van 1806 mensen medewerking: $0{,}69 \times 1806 \approx 1246$ mensen.
Dat is $\frac{1246}{(2849 + 1806)} \cdot 100\% \approx 26{,}8\%$ van de totale steekproef.

6a De mondelinge enquête is niet altijd anoniem en de enquêteur kan door zijn houding, zijn kleding of zijn manier van vragen stellen de antwoorden beïnvloeden.

b

	telefonisch	*schriftelijk*
voordelen	– bij onduidelijke vraag is correctie mogelijk – snel resultaat – men doet sneller mee omdat het weinig tijd vraagt	– is vaak anoniem – in het algemeen royale bedenktijd
nadelen	– niet aselect (geheim nummer of geen telefoon) – kortdurend (weinig bedenktijd) – niet anoniem	– vaak een groot aantal non-response – geen controle op wijze van invullen – invullen is vaak tijdrovend

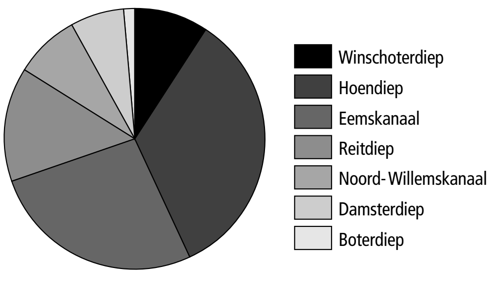
c Ja, want een belangrijke reden om niet mee te doen kan zijn dat het de niet-beantwoorders onverschillig laat terwijl de meerderheid van de overige wijkbewoners sterk voor is in verband met de verkeersveiligheid.

bladzijde 154

7a Percentage schepen door het Winschoterdiep: $\frac{34\,980}{88\,306} \cdot 100\% \approx 39{,}6\%$.

b Bij de schepen, die door het Winschoterdiep gaan hoort een hoek van $0{,}396 \times 360° \approx 142{,}6°$.
Dat stemt overeen met hoek in het cirkeldiagram want die is ongeveer 142°.

c Het totaal van de tonnages is 5 653 151.
Het aandeel van het Winschoterdiep is
$\frac{2\,044\,921}{5\,653\,151} \cdot 100\% \approx 36{,}2\%$.
De bijbehorende hoek in het cirkeldiagram is
$0{,}362 \times 360° \approx 130°$.
Evenzo voor de andere waterwegen:
Hoendiep: $\frac{1\,600\,846}{5\,653\,151} \cdot 360° \approx 102°$
Eemskanaal: $\frac{859\,202}{5\,653\,151} \cdot 360° \approx 55°$
Reitdiep: $\frac{490\,760}{5\,653\,151} \cdot 360° \approx 31°$
Noord-Willemskanaal: $\frac{396\,230}{5\,653\,151} \cdot 360° \approx 25°$
Damsterdiep: $\frac{176\,052}{5\,653\,151} \cdot 360° \approx 11°$ en Boterdiep: $\frac{85\,140}{5\,653\,151} \cdot 360° \approx 5°$.

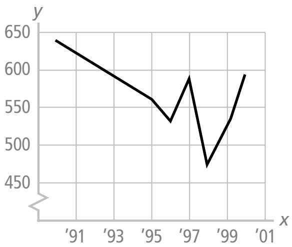


d De sector van het aantal schepen door het Eemskanaal is $0{,}083 \times 360° \approx 30°$.
De sector van het bijbehorende tonnage is ongeveer 55° en die is dus bijna twee keer zo groot.

8a De relatieve frequentie voor het jaar 1990: $\frac{638}{1176} \cdot 100\% \approx 54{,}3\%$.
De relatieve frequenties voor de volgende jaren worden op dezelfde manier berekend.

tijd in jaren	*relatieve frequentie*
1990	54,3%
1995	53,0%
1996	51,9%
1997	56,0%
1998	50,5%
1999	43,25
2000	45,7%

b

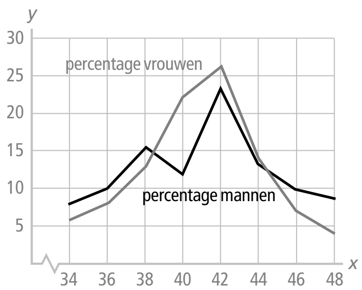


bladzijde 155

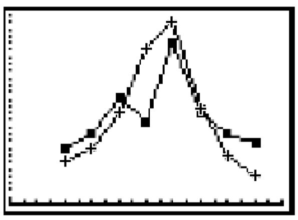
9a De winkelier heeft totaal 151 pantalons op voorraad.
De relatieve frequentie van maat 34: $\frac{12}{151} \cdot 100\% \approx 7{,}9\%$.
De relatieve frequenties van de andere maten worden op dezelfde manier berekend.

maat	34	36	38	40	42	44	46	48
relatieve frequentie van het aantal pantalons	7,9	9,9	15,2	11,9	23,2	13,2	9,9	8,6

b In de tabel staan de percentages vrouwen, die een bepaalde maat kopen.
Deze percentages zijn gelijk aan de bijbehorende relatieve frequenties.



c Vensterinstelling: horizontaal [min, max] = [30, 50]
verticaal [min, max] = [0, 27]



De frequentiepolygoon met de plustekens hoort bij het percentage vrouwen, dat een bepaalde maat koopt.
De frequentiepolygoon met de bloktekens hoort bij de voorraad van de winkelier.

d Daar waar de koopgrafiek boven de voorraadgrafiek ligt is de vraag groter dan het aanbod.
De winkelier doet er daarom verstandig aan om vooral de maten 40 en 42 bij te bestellen.

10a Deze grafiek benadrukt, dat het aantal slachtoffers van spoorwegongelukken sterk is afgenomen.
Het beeld wordt versterkt doordat het lijkt alsof het aantal bijna nul is geworden.
In werkelijkheid is de laagste waarde ruim 2 000 : de verticale as begint bij 2 in plaats van bij 0.

b Ongetwijfeld zal het aantal reizigers in deze periode sterk zijn toegenomen.
Om een uitspraak over de mate van veiligheid te kunnen doen moet ook het totaal aantal reizigers in deze jaren bekend zijn.

c Ja, je kunt het aantal doden in 1932 schatten door de waarde af te lezen die hoort bij het punt van de grafiek, dat recht boven 1932 ligt.
In feite is dit een interpolatie tussen de waarden voor de jaren 1930 en 1935.

d Nee, want de waarden in de grafiek geven het totaal aantal doden voor een heel jaar.
Het is niet bekend hoe dit aantal over de twaalf maanden verdeeld is geweest.

bladzijde 156

11a De volgende lengtes uit de steekproef worden afgerond op 163 cm: 162,9 cm, 163,2 cm en 163,4 cm.

b De linkergrens van deze klasse is 162,5 cm, de rechtergrens van deze klasse is 163,5 cm.

c

klasse	*frequentie*
155 – 159	2
160 – 165	8
165 – 169	6
170 – 174	9

d De linkergrens van de klasse 155 – 159 is 154,5 en de rechtergrens is 159,5.

e De klassenbreedte is $159{,}5 - 154{,}5 = 5$.

f Het klassenmidden van de klasse 155–159 ligt precies in het midden tussen de linker- en de rechterklassengrens en is dus gelijk aan het gemiddelde van beide waarden:
$\frac{154{,}5 + 159{,}5}{2} = 157$.

12a De aantallen 70, 71, 72, 73, 74, 75 en 76; dit zijn alle gehele getallen uit de klasse $[69{,}5\,;76{,}5\rangle$.

b De gewichten vanaf 69,5 tot 76,5 kg, dus alle gewichten uit de klasse $[69{,}5\,;76{,}5\rangle$.

c De klassengrenzen zijn 70 en 77 jaar want iemand wordt pas vanaf z'n 70-ste verjaardag tot de 70-jarigen gerekend en wordt één dag vóór z'n 77-ste verjaardag nog steeds tot de 76-jarigen gerekend.
Het klassenmidden van de klasse $[70\,,77\rangle$ is dus $\frac{70+77}{2} = 73\frac{1}{2}$ jaar.

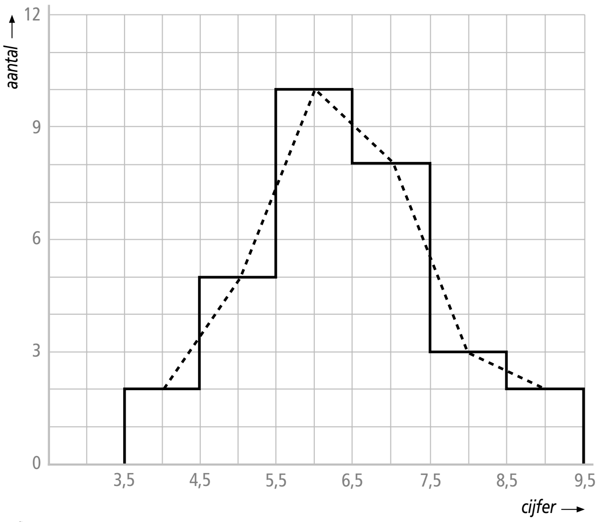
d Theaterkaartjes: klassenmidden is $\frac{69{,}5+76{,}5}{2} = 73$ kaartjes en de klassenbreedte is $76{,}5 - 69{,}5 = 7$.
Gewichten: klassenmidden is $\frac{69{,}5+76{,}5}{2} = 73$ kg en de klassenbreedte is $76{,}5 - 69{,}5 = 7$.
Leeftijden: het klassenmidden is $73\frac{1}{2}$ jaar en de klassenbreedte is $76{,}5 - 69{,}5 = 7$.

bladzijde 157

13a

cijfer	*frequentie*
4	2
5	5
6	10
7	8
8	3
9	2

b



c De klassengrenzen zijn 3,5 en 4,5; in intervalnotatie: $[3{,}5; 4{,}5\rangle$.

d De klassenbreedte is $4{,}5 - 3{,}5 = 1$.

e Zie de figuur bij opdracht b.

f Voordeel staafdiagram: de klassengrenzen zijn vlot af te lezen.
Nadeel staafdiagram: verdeling van de waarnemingen binnen een klasse niet zichtbaar.
Voordeel frequentiepolygoon: waarnemingen binnen een klasse zijn vrij eenvoudig te schatten.
Voordeel frequentiepolygoon: steilste deel geeft grootste stijging of daling ten opzichte van naastgelegen klasse.
Nadeel frequentiepolygoon: de klassengrenzen zijn minder gemakkelijk af te lezen.

14a De klassenbreedte is $2{,}50 \text{ kg} - 2{,}48 \text{ kg} = 0{,}02$ kg.
Het eerste klassenmidden is $\frac{2{,}48 + 2{,}50}{2} = 2{,}49$.
De andere klassenmiddens zijn achtereenvolgens: 2,51, 2,53, 2,55 en 2,57.

b De gewichten van de twee soorten zijn niet goed te vergelijken door gebruik te maken van absolute aantallen omdat het totaal aantal zakken van beide soorten dan (ongeveer) gelijk moet zijn.

c De relatieve frequentie van de zakken Bintjes uit de klasse [2,48; 2,50⟩ is

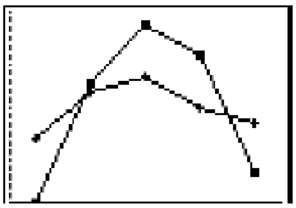
$\frac{4}{30} \cdot 100\% \approx 13,33\%$.

De relatieve frequentie van de zakken Nicola uit de klasse [2,50; 2,52⟩ is

$\frac{12}{45} \cdot 100\% \approx 25,00\%$.

Op dezelfde manier zijn de andere relatieve frequenties te berekenen.

gewicht in kg (klassenmidden)	relatieve frequentie (Bintjes)	relatieve frequentie (Nicola)
2,49	13,33	0,00
2,21	23,33	25,00
2,53	26,67	37,50
2,55	20,00	31,25
2,57	16,67	6,25



d De top van het frequentiepolygoon van de Bintjes ligt niet zo veel hoger dan de
De top van het frequentiepolygoon van de zakken Nicola ligt veel hoger en met name de hoogste en de laagste gewichtsklasse hebben een lage relatieve frequentie.
Samengevat: de gewichten van de zakken Bintjes variëren sterk en de gewichten van de zakken overige punten. Dat betekent, dat er geen grote verschillen zitten in de relatieve frequenties van de gewichtsklassen. Nicola zitten voor het grootste gedeelte in de buurt van 2,53 kg.

15a Het risicoprofiel B (getroffen worden door vallende, schuivende of zwaaiende losse voorwerpen) levert het grootste aantal ongelukken op.

b Voor het aflezen is de volgende regel te gebruiken: 1250 ongelukken hoort bij een staaf, die 40 mm hoog is. Per mm geeft dit dan 31,25 ongelukken.
De totale lengte van alle staven bij elkaar is ongeveer 140 mm.
Dit geeft een totaal aantal ongelukken van ongeveer 4375.

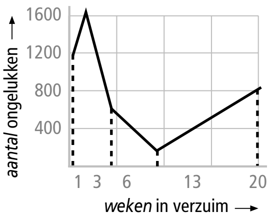
c In de lichtste categorie van ongelukken is de gemiddelde verzuimduur kort en de spreiding is klein.
Bij zwaardere ongelukken is de gemiddelde verzuimduur veel langer en de spreiding is ook groter.
Daarom neemt de klassenbreedte toe naarmate het verzuim langer duurt.

d De totale lengte van alle groene staven bij elkaar is ongeveer 37 mm.
Het aantal ongelukken met minder de één week verzuim is dan ongeveer 1156.
N.B. één mm meer of minder geeft een afwijking van ruim 30 ongelukken meer of minder!

e Met name de klasse 6 – 13 weken is moeilijk af te lezen.
De afgelezen aantallen ongelukken bij elkaar opgeteld leveren een totaal van 4370.
Dit komt goed overeen met het resultaat bij opdracht b.

f

verzuimklasse	aantal ongelukken
< 1 week	1160
1 – 3 weken	1620
3 – 6 weken	630
6 – 13 weken	160
> 13 weken	800



De laatste klasse heeft geen bekende rechtergrens, dus kan het klassenmidden niet worden berekend.
Als we uitgaan van een fictieve klassenbreedte van bijvoorbeeld 14 weken (de rechtergrens is dan 27 weken), dan is het klassenmidden 13 + 7 = 20.
De klassenmiddens van de andere klassen zijn achtereenvolgens: 0,5, 2, 4,5 en 9,5 weken.

g E komt 15 keer zo vaak voor als H.
De gemiddelde verzuimduur is bij E echter veel lager dan bij H, want bij E is 70% na 3 weken weer op het werk terwijl bij H meer als 50% na 13 weken nog steeds niet aan het werk is.
Als een profiel gevaarlijker genoemd wordt als het verzuim gemiddeld langer duurt, dan heeft de directeur dus gelijk.

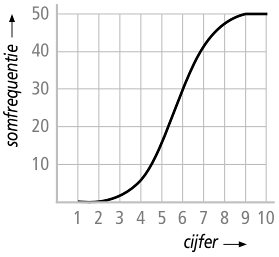
bladzijde 158

16a $2+3+10=15$ leerlingen hebben als eindcijfer hoogstens een 5.

b eerste manier:
tel de frequenties van alle eindcijfers tot en met 7 op.
$2+3+10+16+11=42$.
tweede manier:
tel bij het aantal leerlingen met een eindcijfer hoogstens 5 (is al bekend) de aantallen op die horen bij de leerlingen met eindcijfer 6 en 7.
$15+(16+11)=42$.
derde manier:
trek van het totaal aantal leerlingen het aantal leerlingen af met een eindcijfer hoger dan 7.
50 – (5 + 3) = 42.

c

eindcijfer	somfrequentie
1	0
2	0
3	2
4	5
5	15
6	31
7	42
8	47
9	50
10	50

d 

e De grafiek die bij opdracht d is gemaakt kan niet dalen omdat bij somfrequenties elk volgend getal gelijk is aan het voorgaande getal of groter dan het voorgaande getal.

17a Van de laagste klasse is alleen de rechtergrens af te lezen: 145,5.
De volgende klassen zijn: [145,5; 150,5⟩, [150,5; 155,5⟩, [155,5; 160,5⟩, [160,5; 165,5⟩, [165,5; 170,5⟩, [170,5; 175,5⟩ en [175,5; 180,5⟩.

b Er is 1 meisje met lengte kleiner dan 145,5 cm.
Er zijn $5-1=4$ meisjes met lengte uit de klasse [145,5; 150,5⟩.
Totaal zijn er dus 5 meisjes met een lengte kleiner dan 150,5 cm.
Het aantal 5 wordt dus door een punt van de grafiek aangegeven boven het getal 150,5.
Een soortgelijke redenering geldt ook voor alle andere punten van de grafiek.

c Uit de grafiek: 41 meisjes hadden een lengte van minder dan 160,5 cm.

d 19 – 5 = 14 meisjes hadden een lengte van minstens 150,5 cm en minder dan 155,5 cm.

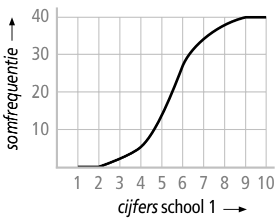
e 100 – 87 = 13 meisjes waren minstens 170,5 cm lang.

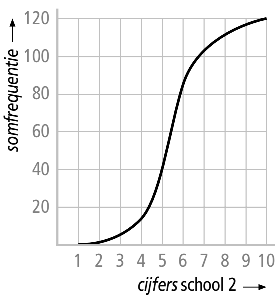
f De klasse [160,5 ; 165,5⟩ bevat het grootste aantal meisjes: 69 – 41 = 28.

bladzijde 159

18a

eindcijfer	*somfrequentie (School 1)*	*somfrequentie (School 2)*
1	0	0
2	0	1
3	2	4
4	5	13
5	14	37
6	27	86
7	34	105
8	38	115
9	40	119
10	40	120

b 



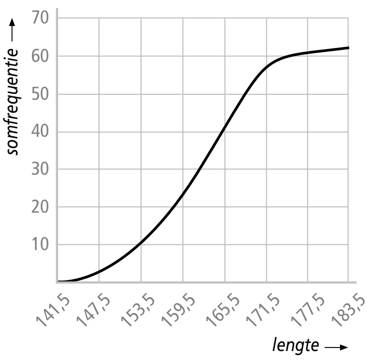
c Het steilste deel van de somfrequentiepolygoon van school 2 gaat van punt $(5,5\,;37)$ naar het punt $(6,5\,;86)$. Dat betekent, dat bij deze klasse de grootste toename van de somfrequentie hoort, dus cijfer 6 komt van alle eindcijfers het meest vaak voor:
86 – 37 = 49 leerlingen hebben eindcijfer 6.

d Het steilste deel van de somfrequentiepolygoon van school 1 gaat van punt $(5{,}5\,;14)$ naar het punt $(6{,}5\,;27)$. De helling van het steilste gedeelte van de somfrequentiepolygoon is bij school 1 kleiner dan bij school 2 omdat de toename van de somfrequentie hier slechts 13 is terwijl dat bij school 2 bij dezelfde klassenbreedte 49 is.

19a

lengte in klassen	*frequentie*	*somfrequentie*
[141,5 ; 147,5⟩	2	2
[147,5 ; 153,5⟩	7	9
[153,5 ; 159,5⟩	12	21
[159,5 ; 165,5⟩	20	41
[165,5 ; 171,5⟩	17	58
[171,5 ; 177,5⟩	3	61
[177,5 ; 183,5⟩	1	62

b



c In de somfrequentietabel zal van de somfrequenties alleen het laatste getal veranderen, dat zal 3 groter worden.
Daardoor zal het laatste punt van de somfrequentiepolygoon veranderen van $(183{,}5\,;62)$ in $(183{,}5\,;65)$ en het laatste deel van de polygoon zal iets steiler lopen.

d In de somfrequentietabel worden de getallen in de kolom somfrequentie met 3 verhoogd.
Alle punten van de polygoon, met uitzondering van het eerste punt $(141{,}5\,;0)$ komen 3 eenheden hoger te liggen.
Het eerste deel van de polygoon zal daardoor iets steiler lopen, alle volgende delen hebben dezelfde helling als die van de oorspronkelijke somfrequentiepolygoon.

bladzijde 160

20a 196 mensen gingen naar 3 concerten.

b Er zijn totaal $375 \cdot 1 + 293 \cdot 2 + 196 \cdot 3 + 163 \cdot 4 = 2201$ kaarten verkocht.

c 1027 mensen hebben in totaal 2201 kaarten gekocht.
Het gemiddeld aantal concertbezoeken per persoon is $\frac{2201}{1027} \approx 2{,}14$.

bladzijde 161

21a Zowel in klas 2A als in klas 2B zitten 24 leerlingen.

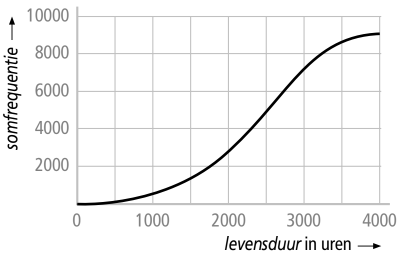
b Het cijfer 5 komt het meeste voor, de modus is dus 5.
De mediaan is het gemiddelde van het twaalfde cijfer (= 5) en het dertiende cijfer (= 6) als de cijfers worden gerangschikt van klein naar groot: $\frac{1}{2}(5+6)=5{,}5$.
Het gemiddelde is $\frac{1\cdot 3+3\cdot 4+8\cdot 5+3\cdot 6+3\cdot 7+3\cdot 8+2\cdot 9+1\cdot 10}{24}=\frac{146}{24}\approx 6{,}1$.

c Het cijfer 6 komt het meeste voor, de modus is dus 6.
De mediaan is het gemiddelde van het twaalfde cijfer (= 6) en het dertiende cijfer (= 6) als de cijfers worden gerangschikt van klein naar groot: $\frac{1}{2}(6+6)=6$.
Het gemiddelde is $\frac{3\cdot 2+4\cdot 3+4\cdot 4+10\cdot 6+2\cdot 7+1\cdot 8}{24}=\frac{116}{24}\approx 4{,}8$.

d De cijfers en de frequenties worden ingevoerd in de rekenmachine (eerste scherm / TI-84).
Het tweede scherm verschijnt na de commando's STAT / CALC / 1-Var Stats L_1, L_2.
Bovenaan staat het gemiddelde. Door het tweede scherm verder naar beneden te scrollen ontstaat het derde scherm. Hierin is de mediaan af te lezen.

e Het aantal onvoldoendes / voldoendes is in beide klassen ongeveer gelijk verdeeld.
Het gemiddelde van klas 2A is hoger dan van klas 2B terwijl de andere twee centrummaten voor klas 2B hogere waarden geven dan voor klas 2A.
In dit geval geeft het gemiddelde de verschillen tussen de klassen dus het beste weer.

22a



b Het steilste gedeelte van de somfrequentiepolygoon ligt tussen de punten $(2000, 2687)$ en $(2500, 5033)$.
De modale klasse is dus de klasse [2000, 2500⟩.

c

klassenmidden	*frequentie*
250	45
750	355
1250	753
1750	1534
2250	2346
2750	2287
3250	1458
3750	344

De frequentie van de klasse [500, 1000⟩ is gelijk aan 400 – 45 = 355.
Op dezelfde manier worden de frequenties van de overige klassen berekend (zie tabel).
De gemiddelde levensduur wordt als volgt benaderd:

$$\frac{45\cdot 250+355\cdot 750+\ldots\ldots+344\cdot 3750}{9122}\approx 2357 \text{ uur.}$$

d $\frac{9122}{2} = 4561$, dus de mediaan is het gemiddelde het 4561^e en 4562^e getal als alle getallen worden gerangschikt van klein naar groot.
De mediaan bevindt zich in de klasse $[2\,000, 2\,500\rangle$.
Schatting met behulp van de somfrequentiepolygoon: geef op de verticale as bij benadering de plaats van het getal 4561,5 aan, trek een horizontale lijn naar de polygoon en vanaf het snijpunt een verticale lijn.
Het snijpunt van deze verticale lijn met de horizontale as is een benadering van de mediaan: ongeveer 2400.
Schatting met behulp van een berekening: $2000 + \frac{4561{,}5 - 2687}{5033 - 2687} \cdot (2500 - 2000) \approx 2400$.

23a $\frac{1500}{2} = 750$. De mediaan ligt dus in de klasse $[0{,}10\,000\rangle$.
De modale klasse is $[0{,}10\,000\rangle$.

b Het salaris van de directeur wordt op € 12 000 gesteld.
Schatting van het gemiddelde salaris:

$$\frac{752 \cdot 5\,000 + 663 \cdot 17\,500 + 76 \cdot 32\,500 + 8 \cdot 50\,000 + 1 \cdot 120\,000}{1500} = 12\,235 \text{ euro.}$$

c De mediaan ligt dus in de klasse $[0, 10\,000\rangle$ en de modale klasse is $[0, 10\,000\rangle$ terwijl het gemiddelde in de klasse $[10\,000, 25\,000\rangle$ ligt.
De eigenaar zal dus het gemiddelde gebruiken.

d Het gegeven dat de mediaan in de laagste klasse ligt benadrukt dat meer dan de helft van de werknemers een salaris heeft dat lager is dan 10 000 euro.

bladzijde 162

24a Rechts van het getal 18 komt driemaal het getal 5 voor.

b Het gaat in totaal om 18 flesjes.
De mediaan is het gemiddelde van het 9^e en het 10^e getal als de getallen van klein naar groot zijn gerangschikt.
De mediaan is 185 cl want zowel het 9^e als het 10^e getal betreft een flesje met een inhoud van 185 cl.

c De eerste groep bestaat uit 9 waarnemingsgetallen.
De mediaan daarvan is het 5^e getal: 182 cl.

d Van 4 flesjes is de inhoud kleiner dan 182 cl: $\frac{4}{18} \cdot 100\% \approx 22{,}2\%$.

25a 5 uur en 10 minuten valt samen met het derde kwartiel: 75%.

b Het derde kwartiel bij de mannen valt samen met het eerste kwartiel bij de vrouwen: 75%.

c De mediaan bij de vrouwen valt samen met een tijd van 5 uur en 25 minuten.
De snelste vrouw was al na 4 uur en 30 minuten binnen.
De snelste 50% zit dus tussen 4 uur en 30 minuten en 5 uur en 25 minuten.

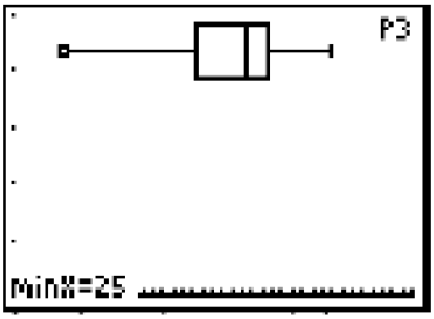
d Tussen de mediaan en het derde kwartiel zit 50%.

e De eerste man ging na 4 uur over de streep, de laatste na 5 uur en 30 minuten.
Daar zit dus 1 uur en 30 minuten tussen.

f De eerste vrouw ging na 4 uur en 30 minuten over de streep, de laatste na 6 uur.
Het verschil is dus 1 uur en 30 minuten.

bladzijde 163

26a



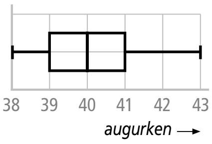
De afstand tussen de hoogste en de laagste score is 52 – 25 = 27

b Het eerste kwartiel is uit de boxplot af te lezen: 38,5.
Evenzo het derde kwartiel: 45,5.

c De mediaan is 43,5.
De afstand tussen Q_1 en mediaan is 43,5 – 38,5 = 5.
De afstand tussen Q_3 en mediaan is 45,5 – 43,5 = 2.
Zowel tussen Q_1 en mediaan als tussen Q_3 en mediaan zit 25% van de scores, maar tussen Q_3 en mediaan zitten de scores blijkbaar dichter bij elkaar.

27a Er zijn 8 blikken gevonden waarin 38 augurken zijn geteld
Het getal 38 is dus een waarneming en het getal 8 is de bijbehorende frequentie.

b Voer de gegevens van de tabel in je rekenmachine in en maak een plot van de boxplot. Die komt dan overeen met de figuur hieronder.

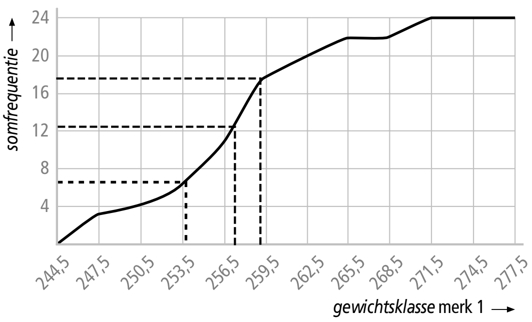


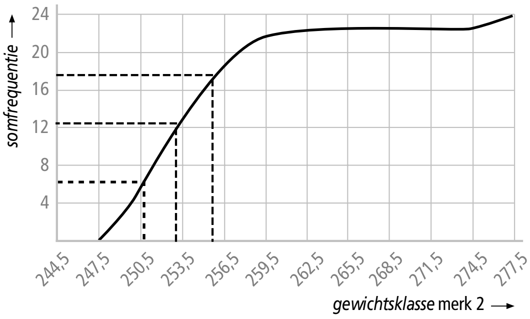
De mediaan en de kwartielen kunnen ook als volgt worden berekend.
Er zijn in totaal 119 blikken. De mediaan is het 60ᵉ waarnemingsgetal: 40.
De mediaan wordt nu weggelaten; er blijven twee groepen van 59 waarnemingen over.
Het eerste kwartiel is het 30ᵉ waarnemingsgetal: 39.
Het derde kwartiel is het 90ᵉ waarnemingsgetal: 41.

c Het kleinste aantal augurken is 38 en het grootste aantal is 43.
De spreidingsbreedte is dus 43 – 38 = 5.

d De kwartielafstand is $Q_3 - Q_1 = 41 - 39 = 2$.

28a





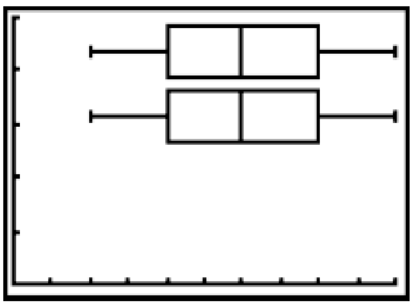
b Uit beide polygonen zijn bij benadering onderstaande waarden af te lezen:

	merk 1	*merk 2*
Q_1	254	251
mediaan	257	253
Q_3	259	256

c De spreidingsmaat geeft vooral voor merk 2 een vertekend beeld omdat de waarde hiervan extra groot uitvalt door slechts 1 pak dat relatief zwaar is.
Dit bezwaar geldt in iets mindere mate ook voor merk 1: het laatste deel van de polygoon is erg vlak.
Voor beide merken is de kwartielafstand de beste spreidingsmaat.

bladzijde 164

29a In de plot hieronder is te zien, dat voor Jacolien en Gijs de spreidingsbreedte en de mediaan precies gelijk zijn.
Verder valt op dat voor beiden de mediaan gelijk is.

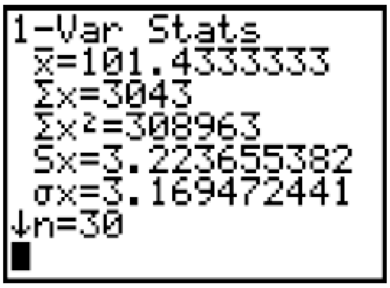


b Het gemiddelde van Jacolien: $\dfrac{2+2\cdot 4+4\cdot 6+2\cdot 8+10}{10}=6.$
Het gemiddelde van Gijs: $\dfrac{2\cdot 2+3\cdot 4+3\cdot 8+2\cdot 10}{10}=6.$

c Gijs heeft vijf diepe onvoldoendes en daarnaast ook vijf hoge cijfers.
De cijfers van Jacolien zijn wat gelijkmatiger verdeeld:
drie diepe onvoldoendes, vier zessen en drie hoge cijfers.

30a 9 van de 30 pakjes hebben ondergewicht: $\frac{9}{30} \cdot 100\% = 30\%$.

b Met de rekenmachine wordt het gemiddelde gewicht en de standaarddeviatie berekend (zie plot hieronder): het gemiddelde gewicht is ongeveer 101,4 gram en de standaarddeviatie is ongeveer 3,2 gram.



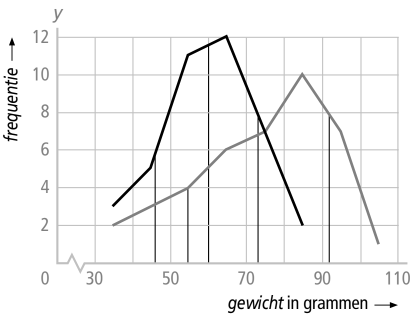
bladzijde 165

31a Met behulp van de rekenmachine (gewichten in L_1 en bijbehorende frequenties in L_2) worden het gemiddelde en de standaardafwijking berekend:het gemiddelde is ongeveer 101,4 en de standaardafwijking is ongeveer 2,31.

b 7 van de 48 pakjes hebben ondergewicht: $\frac{7}{48} \cdot 100\% \approx 14,6\%$.

c Het percentage pakjes met ondergewicht is kleiner.
Dit wordt veroorzaakt door het feit dat bij hetzelfde gemiddelde de standaardafwijking kleiner is.

32a De frequentiepolygoon van boom A is de vette grafiek in onderstaande figuur.



b Met behulp van de rekenmachine (gewichten in L_1 en bijbehorende frequenties in L_2) worden het gemiddelde en de standaardafwijking berekend: het gemiddelde is ongeveer 59,6 en de standaardafwijking is ongeveer 13,4.

c Het gemiddelde min standaardafwijking: $59,55 - 13,42 \approx 46,1$.
Het gemiddelde plus standaardafwijking: $59,55 + 13,42 \approx 73,0$.

d In het interval [46,1; 73,0] liggen 26 waarnemingen: $\frac{26}{40} \cdot 100\% = 65\%$.

e Het frequentiepolygoon van boom B is in dezelfde figuur getekend als het frequentiepolygoon van boom A (de groene grafiek).
Het bijbehorende gemiddelde de bijbehorende standaarddeviatie zijn achtereenvolgens ongeveer 73,4 en 18,8.

Het gemiddelde min standaardafwijking: $73,4-18,8=54,6$.
Het gemiddelde plus standaardafwijking: $73,4+18,8\approx 92,2$.
In het interval [54,6; 92,2] liggen 26 waarnemingen: $\frac{26}{40}\cdot 100\% = 65\%$.

f Boom B staat vermoedelijk in een park omdat het gemiddeld gewicht van de kastanjes door de gunstigere omstandigheden hoger is.
Dat is ook al in het diagram te zien omdat er bij boom B meer kastanjes zijn in de hogere gewichtsklassen.

33a Het gemiddelde min de standaarddeviatie: $99,99-0,15=99,84$.
Het gemiddelde plus de standaarddeviatie: $99,99+0,15=100,14$.
In het interval $[99,84\,;100,14]$ liggen 272 waarnemingen: $\frac{272}{400}\cdot 100\% = 68\%$.

b Afgekeurd worden 7 + 21 + 6 + 3 = 37 spaken: $\frac{37}{400}\cdot 100\% \approx 9\%$.

bladzijde 166

34a Met behulp van de rekenmachine (klassenmiddens in L_1 en bijbehorende frequenties in L_2) worden het gemiddelde en de standaardafwijking berekend: het gemiddelde is 2064 en de standaardafwijking is ongeveer 391,29.
Afgerond op € 100,- is dat achtereenvolgens € 2100,- en € 400,-.

b Het salaris bij de vrouwen ligt gemiddeld lager want bijna de helft van de vrouwen heeft een salaris in de groepen 1 t/m 5 en bij de mannen is dat 30%.

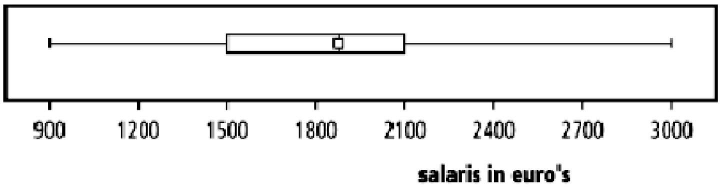
c Er moet nu eerst een nieuwe tabel worden gemaakt met daarin per salarisgroep het totaal aantal werknemers; de twee kolommen voor mannen en vrouwen afzonderlijk worden dus bij elkaar opgeteld.

salarisgroep	*klassenmidden*	*frequentie (vrouwen + mannen)*
1	900	2
2	1100	6
3	1300	11
4	1500	9
5	1700	22
6	1900	20
7	2100	21
8	2300	16
9	2500	10
10	2700	6
11	2900	1
12	3100	1

Met behulp van de rekenmachine (klassenmiddens in L_1 en bijbehorende frequenties in L_2) wordt het gemiddelde berekend: 1917,6.
Afgerond op € 100,- is dat € 1900,-.

d De mediaan is het $\frac{75+1}{2} = 38^e$ waarnemingsgetal: 1900.
Het eerste kwartiel is het $\frac{37+1}{2} = 19^e$ waarnemingsgetal: 1500.

Het derde kwartiel is het 38 + 19 = 57^e waarnemingsgetal: 2100.



e Bij de mannen ligt de mediaan bij € 2100,- , dus 50% van de mannen verdient meer dan € 2100,-. Bij de vrouwen ligt het derde kwartiel bij € 2100,- , dus 25% van de vrouwen verdient meer dan € 2100,-.
Naar verhouding zijn er dus tweemaal zo veel mannen als vrouwen, die meer dan € 2100,- verdienen.
Als we naar de absolute aantallen kijken, dan klopt deze verhouding 2:1 niet.

35a Aflezen bij 2500 gram geeft: 18%.
b Aflezen bij 3000 gram geeft 39% en aflezen bij 3500 gram geeft 66%.
Het percentage kinderen met een geboortegewicht tussen de 3000 en 3500 gram is dus $66\% - 39\% = 27\%$.
c Mediaan: aflezen bij een percentage van 50 geeft een gewicht van ongeveer 3200 gram.
d Aflezen bij een percentage van 20 geeft een gewicht van ongeveer 2650 gram: $P_{20} \approx 2650$.
e $(100 - 95)\% = 5\%$ van de kinderen heeft een gewicht hoger dan P_{95} en 5% van de kinderen heeft een gewicht lager dan P_5.
Van de 8428 kinderen had dus $(5+5)\% = 10\% \approx 843$ kinderen een gewicht hoger dan P_{95} of een gewicht lager dan P_5.

bladzijde 167

36a Het klassenmidden van de klasse met 22, 23 en 24 is 23.
Aflezen bij 23 geeft: 6% van de zinnen uit de BZ bevatten 22, 23 of 24 woorden. (zie ook de frequentietabel).
b Het laatste punt van de grafiek van de BZ is (29, 1).
Dat betekent, dat 1% van de zinnen tot de klasse [28, 30] behoort.
Zinnen van 30 woorden komen dus voor, maar zinnen van 31 woorden niet.
c De grafieken van de FAZ en de WT snijden elkaar in de punten (8, 4) en (26, 9).
Dat betekent, dat zinnen met een lengte van 7, 8 of 9 woorden in de FAZ en in de WT dezelfde frequentie hebben, namelijk 4%.
Datzelfde geldt voor zinnen met een lengte van 25, 26 of 27 woorden met een frequentie van 9%.
d In de onderzochte teksten van de FAZ kwamen geen zinnen voor met een lengte van minder dan 7 woorden en in de WT wel.

e De berekende gegevens staan in onderstaande tabel.

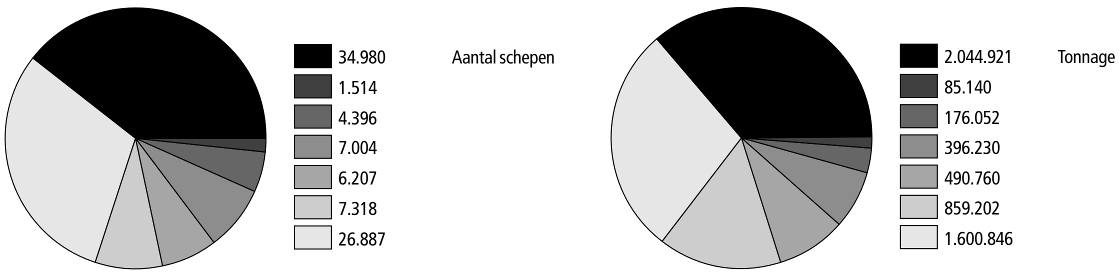
	BZ	FAZ	WT
modus (modale klasse)	[10, 12]	geen *)	[19, 21]
mediaan in de klasse	[10, 12]	[22, 24]	[25, 27]
gemiddelde(benadering)	13,3	24,2	26,5

De verschillen worden zowel door de mediaan als door het gemiddelde goed in beeld gebracht.
*) Er zijn twee klassen met dezelfde (hoogste) frequentie.

f De wetenschappelijke tekst is het moeilijkst leesbaar want lange zinnen komen daar het meeste voor en de gemiddelde lengte van de zinnen is in de WT het grootst.

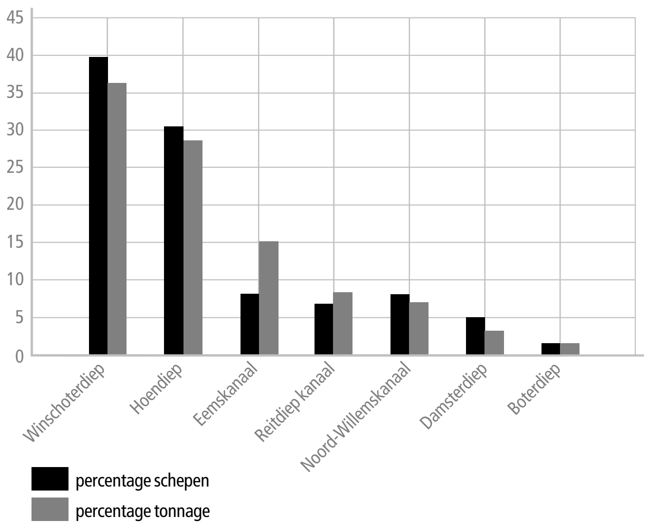
bladzijde 168

I-1a

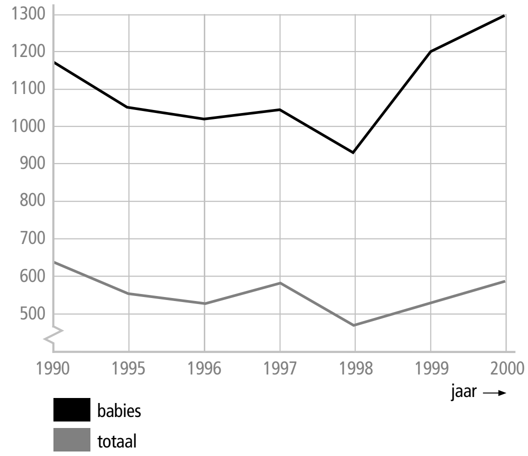


b In het cirkeldiagram van het aantal schepen is de cirkelsector groter dan de cirkelsector in het cirkeldiagram van het tonnage.

c



Bij het Eemskanaal is het percentage van het tonnage ten opzichte van het percentage van het aantal schepen veel groter dan bij de andere waterwegen. Dat komt omdat er door het Eemskanaal grotere schepen met een grotere laadruimte kunnen varen. Per schip wordt er dus gemiddeld meer lading vervoerd.

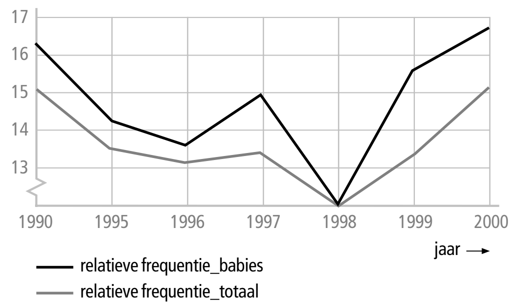
I-2a 

b In de periode 1998 – 1999 is de toename van het totaal aantal adopties groter dan de toename van het aantal adopties van baby's.

c

jaar	*relatieve frequentie baby's*	*relatieve frequentie totaal*
1990	16,33	15,16
1995	14,31	13,60
1996	13,64	13,24
1997	15,02	13,51
1998	12,11	12,08
1999	13,41	15,65
2000	15,18	16,75

d Van alle geadopteerde baby's gedurende deze zeven jaren (3 907) is 16,33% in het jaar 1990 geadopteerd.

e 

Het grillige verloop wordt versterkt door het feit, dat de verticale as pas bij 12 begint.
Als de verticale as bij 0 begint, dan blijkt dat het verloop helemaal niet zo grillig is.

bladzijde 169

I-3a Voor het berekenen van een percentage gebruik je de formule: $\frac{\text{deel}}{\text{geheel}} \cdot 100\%$.
Voorbeeld voor het jaar 1990: $\frac{638}{1176} \cdot 100\% \approx 54{,}25\%$.

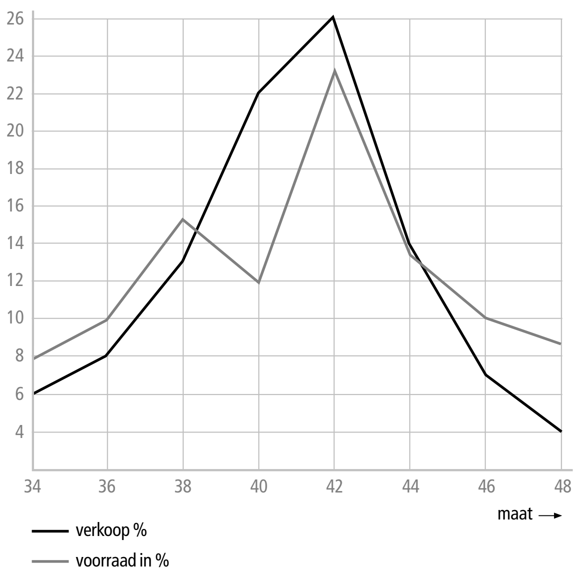
b

jaar	*baby's*	*totaal*	*baby's in % van het totaal*
1990	638	1176	54,25
1995	559	1055	52,99
1996	533	1027	51,90
1997	587	1048	56,01
1998	473	937	50,48
1999	524	1214	43,16
2000	593	1299	45,65

c

d Het aandeel van de baby's in het totaal van de geadopteerde kinderen is in de loop der jaren kleiner geworden.
Naar verhouding worden er in de latere jaren meer oudere kinderen geadopteerd.

I-4a



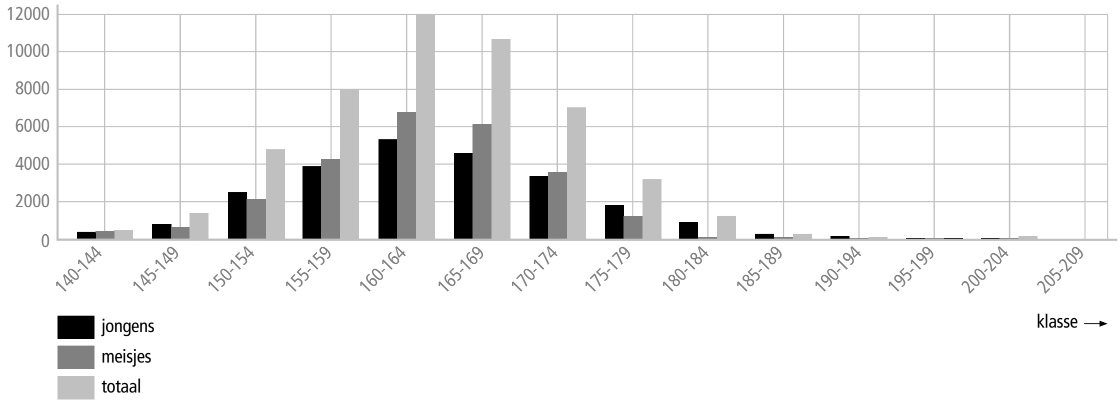
De voorraad en de verkoop zijn aardig met elkaar in overeenstemming met uitzondering van de voorraad van maat 40.

b De winkelier krijgt het advies om in ieder geval maat 40 te bestellen.
Ook de maten 42 en 44 hebben enige aanvulling nodig.

I-5a Het verloop van de grafiek suggereert, dat het aantal doden in de VS door spoorwegongelukken sterk gedaald is. de suggestie van het dalend effect is versterkt door de verticale as niet met 0 maar met 2 te laten beginnen.

b In de loop der jaren is het aantal reizigers per trein en het aantal treinen enorm toegenomen.
Om een goed beeld van de daling te krijgen moeten we dus ook over deze data beschikken.

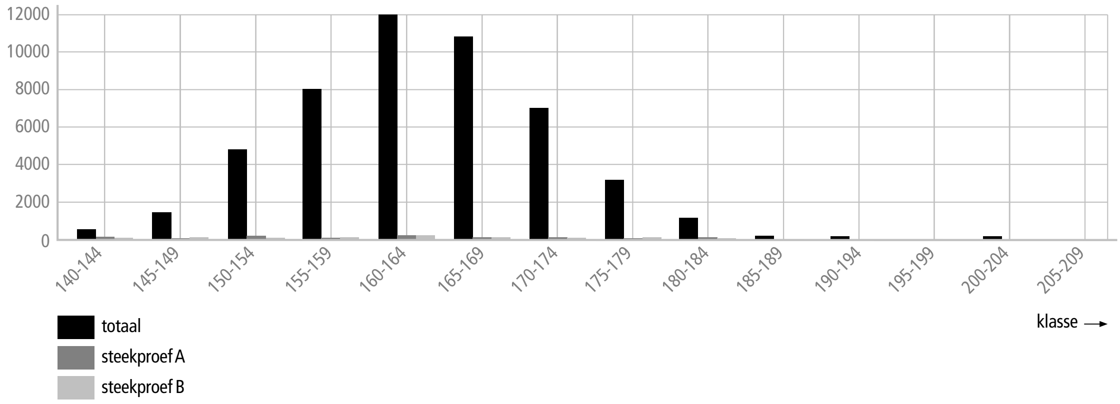
I-6a



De hoogte van de staaf ‘totaal’ is gelijk aan de som van de hoogten van de staven ‘jongens’ en ‘meisjes’.

b Het totaal aantal meisjes is ongeveer 5% groter dan het totaal aantal jongens.
Toch is er met name bij de klassen [160, 164] en [165, 169] een aanzienlijk groter aantal meisjes. Bij de klassen met de grootste lengtes zijn de jongens in de meerderheid.
Dit geldt ook voor de klassen met de kleinste lengtes.
Bij de jongens is de spreiding van de lengte dus groter dan bij de meisjes.
Dat heeft te maken met het feit, dat de start van de ‘groeispurt’ bij de jongens een grotere spreiding vertoont dan de bij de meisjes.

c

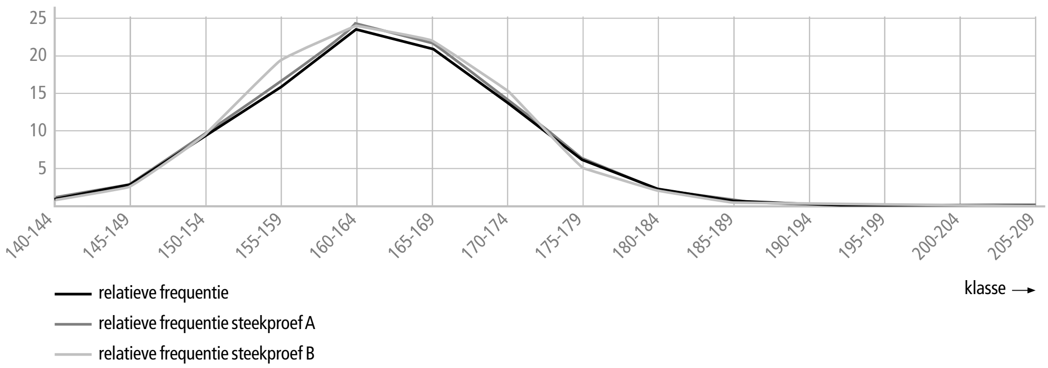


De lengtes van de staven van de steekproeven A en B vallen in het niet bij de lengtes van de staven van het totaal.
De staven van de steekproeven A en B zijn dus nauwelijks te zien en daarom kan er geen goede vergelijking gemaakt worden.

d Van de drie frequentietabellen wordt een tabel met de relatieve frequenties gemaakt.

klasse	*rel. frequentie totaal*	*rel. frequentie steekproef A*	*rel. frequentie steekproef B*
140-144	1.09	1.40	0.80
145-149	2.97	3.00	2.60
150-154	9.64	9.60	9.60
155-159	16.40	16.00	17.60
160-164	24.30	24.60	24.00
165-169	21.74	21.60	22.00
170-174	14.23	13.80	15.40
175-179	6.28	5.80	5.00
180-184	2.31	3.00	2.00
185-189	0.67	1.20	0.40
190-194	0.21	0.00	0.20
195-199	0.07	0.00	0.20
200-204	0.08	0.00	0.20
205-209	0.00	0.00	0.00

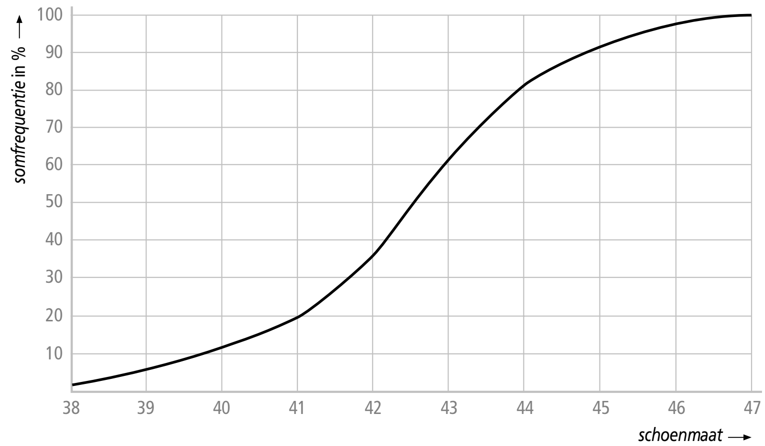
Daarvan worden vervolgens in één figuur relatieve frequentiepolygonen gemaakt.
In de figuur hieronder is te zien, dat deze polygonen vrijwel samenvallen.
De beide steekproeven geven dus een goed beeld van de 50 000 data.



bladzijde 172

T-1a

schoenmaat	*frequentie*	*somfrequentie in %*
38	1	2
39	2	6
40	3	12
41	4	20
42	8	36
43	13	62
44	10	82
45	5	92
46	3	98
48	1	100



T-2a Partij 1 bestaat uit 19 kiwi's. De mediaan is het gewicht van de 10^e kiwi: 77 gram.
Partij 1 bestaat uit 18 kiwi's. De mediaan is het gemiddelde gewicht van de 9^e en de 10^e kiwi: beide kiwi's wegen 80 gram, dus de mediaan is 80 gram.

b De modus is het waarnemingsgetal, dat het meest voorkomt.
Bij partij 1 is dat 77 gram (3 keer) en bij partij 2 is dat 79 gram (3 keer).

c Het gemiddelde gewicht van partij 1: $\frac{1501}{19} = 79$ gram.

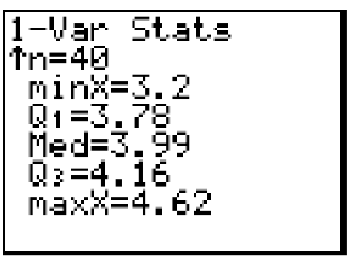
Het gemiddelde gewicht van partij 2: $\frac{1441}{18} \approx 80{,}1$ gram.

d Bij partij 2 komen mediaan en gemiddelde vrijwel overeen.
Bij partij 1 wordt het gemiddelde wat naar boven getrokken door vooral de afwijkende waarden 89 en 91 gram. Hier is het gemiddelde 2 groter dan de mediaan.
De modus is niet zo geschikt omdat de meeste waarnemingsgetallen frequentie 1 hebben.
De mediaan is dus het meest geschikt om beide partijen te vergelijken.

T-3a Waarneming 4,06% komt het meest voor (4 keer): modus is 4,06%.
Er zijn 40 waarnemingen. De mediaan is het gemiddelde van de 20^e en de 21^e waarneming:
het gemiddelde van 3,98 en 4,00 is 3,99%.

Het gemiddelde van alle waarnemingen is $\frac{159{,}08}{40} \approx 3{,}98$ %.

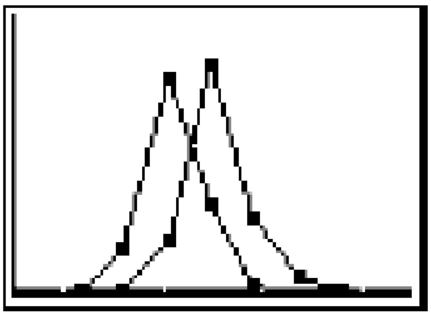
Er zijn 20 waarnemingen kleiner dan de mediaan.
Q_1 is het gemiddelde van de 10^e en de 11^e waarneming: 3,78%.
Er zijn 20 waarnemingen groter dan de mediaan.
Q_3 is het gemiddelde van de 30^e en de 31^e waarneming: 4,16%.
Bovenstaande uitkomsten zijn ook te vinden met behulp van de grafische rekenmachine.



b De spreidingsbreedte is $4{,}62\% - 3{,}20\% = 1{,}42\%$.
De kwartielafstand is $Q_3 - Q_1 = 4{,}16\% - 3{,}785 = 0{,}38\%$.
c 4,05 – 3,70 = 0,35%: de kwartielafstand is dus niet toegenomen.
d Q_1 van het voorafgaande jaar is 3,78%.
Het laagste vetgehalte een jaar later is 3,40%, en dat is dus niet groter dan 25% van de vetgehaltes in het voorafgaande jaar.
e Van de waarnemingen een jaar later is de mediaan 3,85%: dat is aanzienlijk kleiner dan 3,99%.
Van de waarnemingen een jaar later is Q_3 = 4,05%: dat is aanzienlijk kleiner dan 4,16%.
Deze en andere argumenten leiden tot de conclusie, dat het vetgehalte in een jaar niet zal zijn toegenomen.

bladzijde 173

T-4a



De (rechter-)grafiek in de figuur hierboven is bij benadering symmetrisch.
b De frequentiepolygoon van de meisjes is de linkergrafiek in de figuur hierboven.
Ook deze grafiek is bij benadering symmetrisch.
c Het is niet bekend hoe de waarnemingen binnen een klasse zijn verdeeld omdat de oorspronkelijke data niet gegeven zijn. Door gebruik te maken van de klassenmiddens is het gemiddelde en de standaarddeviatie te benaderen. Voer de klassenmiddens en de bijbehorende frequenties in drie afzonderlijke lijsten in de grafische rekenmachine in en bereken het gemiddelde en de standaarddeviatie:
Meisjes: gemiddelde 66,8 cm en standaardafwijking 6,9 cm
Jongens: gemiddelde 76,4 cm en standaardafwijking 8,6 cm
d Jongens: [gemiddelde – SD, gemiddelde + SD] = $[76{,}4 - 8{,}6; 76{,}4 + 8{,}6] = [67{,}8; 85{,}0]$.
In de klasse $[60, 70\rangle$ zitten 29 waarnemingen.
Hiervan worden $\frac{7{,}8}{10} \cdot 29 \approx 23$ kleiner dan 67,8 gerekend.
De resterende 6 waarnemingen worden dus tot het interval $[67{,}8; 70\rangle$ gerekend.
In de klasse $[80, 90\rangle$ wordt de helft van de waarnemingen kleiner dan 85,0 gerekend.
Dat zijn ongeveer 20 waarnemingen.
Een schatting van het aantal waarnemingen binnen het interval $[67{,}8; 85{,}0]$ levert een totaal van $6 + 123 + 20 = 149$ jongens en dat komt overeen met $\frac{149}{209} \cdot 100 \approx 71{,}3\%$.
Meisjes: [gemiddelde – SD, gemiddelde + SD] = $[66{,}8 - 6{,}9; 66{,}8 + 6{,}9] = [59{,}9; 73{,}7]$.
In de klasse $[50, 60\rangle$ zitten 25 waarnemingen.
Deze waarnemingen worden allemaal kleiner dan 59,9 gerekend.
In de klasse $[70, 80\rangle$ zitten 49 waarnemingen.
Hiervan worden er $\frac{3{,}7}{10} \cdot 49 \approx 18$ kleiner dan 73,7 gerekend.

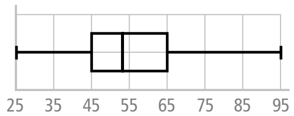
Een schatting van het aantal waarnemingen binnen het interval $[59,9; 73,7]$ levert een totaal van $116+18=134$ meisjes en dat komt overeen met $\frac{134}{195}\cdot 100 \approx 68,7\%$. Het percentage bij de meisjes is inderdaad ongeveer 70%. De conclusie van de onderzoeker dat het percentage bij de jongens ongeveer 80% is blijkt niet juist te zijn.

T-5a Ongeveer 55 leerlingen behaalden een score van meer dan 55 punten.

b Alle 100 leerlingen behaalden een score van maximaal 95 punten, dus geen enkele leerling kan 110 punten voor de test gehaald hebben.

c 25 leerlingen hadden een score van ongeveer 45 of lager. Q_1 is ongeveer 45.
50 leerlingen hadden een score van ongeveer 53 of lager. De mediaan is ongeveer 53.
75 leerlingen hadden een score van ongeveer 63 of lager. Q_3 is ongeveer 63.

d



T-6a In deze straat wonen $6+10+7+3+2+1=29$ gezinnen.

b Het gemiddelde aantal kinderen per gezin:

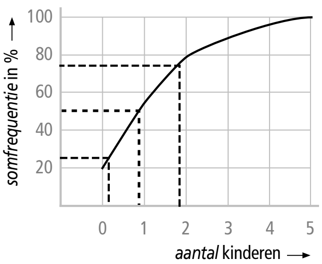
$$\frac{6\times 0+10\times 1+7\times 2+3\times 3+4\times 2+5\times 1}{29}=\frac{46}{29}\approx 1,59 .$$

c In het dorp wonen naar schatting $890\times 1,59\approx 1415$ kinderen.

d De gegevens van de grafiek zijn eerst in onderstaande tabel gezet.
In de derde kolom staat de berekende somfrequentie in %.

aantal kinderen	*frequentie*	*somfrequentie in %*
0	6	20,7
1	10	55,1
2	7	79,3
3	3	89,7
4	2	96,6
5	1	100

Van deze gegevens is hieronder een somfrequentiepolygoon gemaakt.



e In de frequentiepolygoon hierboven is de kwartielafstand af te lezen:
$Q_3 - Q_1 \approx 1,85 - 0,15 = 1,70$.