

2. Data en datasets verwerken

Boekje 2 havo wiskunde A, domein E: Statistiek

Uitwerkingen



Verantwoording



© 2015, SLO (nationaal expertisecentrum leerplanontwikkeling), Enschede

Dit lesmateriaal is ontwikkeld in het kader van de nieuwe examenprogramma's zoals voorgesteld door de commissie Toekomst Wiskunde Onderwijs (cTWO) en herzien door SLO.

Mits de bron wordt vermeld, is het toegestaan zonder voorafgaande toestemming van de uitgever deze uitgave geheel of gedeeltelijk te kopiëren en/of verspreiden en om afgeleid materiaal te maken dat op deze uitgave is gebaseerd.

Auteurs: Erik van Barneveld, Wouter Boer, Carel van de Giessen, Peter Kop, Heleen van der Ree, Henk Reuling, Frits Spijkers, Tanja Stroosma, Anneke Verschut

Met medewerking van: Nico Alink, Martine de Klein (eindredactie)

Informatie: SLO
Afdeling: tweede fase
Postbus 2041, 7500 CA Enschede
Telefoon (053) 4840 661
Internet: www.slo.nl
E-mail: jktweedefase@slo.nl



Inhoud

§ 2.1	Data presenteren.....	4
§ 2.2	Verbanden tussen datarepresentaties.....	13
§ 2.3	Frequentieverdelingen typeren.....	16
§ 2.4	Twee groepen vergelijken	23
§ 2.5	Samenhang tussen twee variabelen	29



§ 2.1 Data presenteren

📄 Opgave 1

- Geslacht, geboortemaand, geboortemaand, gewicht, lengte, gemiddeld cijfer over alle vakken, aantal uren huiswerk per week, wiskunde A of B, gekozen profiel, met plezier naar school.
- Geslacht.

📄 Opgave 2

- ICT.
- Wouters gewicht is redelijk hoog; zijn cijfergemiddelde is laag t.o.v. de rest van de groep.

📄 Opgave 3

- ICT.
- Zie vraag 2b.

📄 Opgave 4

- ICT.
- Zie vraag 2b; je weet nu echter niet meer precies hoeveel anderen er meer wegen of een lager cijfer hebben.

📄 Opgave 5

- ICT.
- Zie vraag 4b.

📄 Opgave 6

- ICT.
- Zie vraag 2b.

📄 Opgave 7

- ICT.
- Je ziet meteen ... Zie vraag 2b.

📄 Opgave 8

- ICT.
- Je ziet meteen ... Zie vraag 2b.

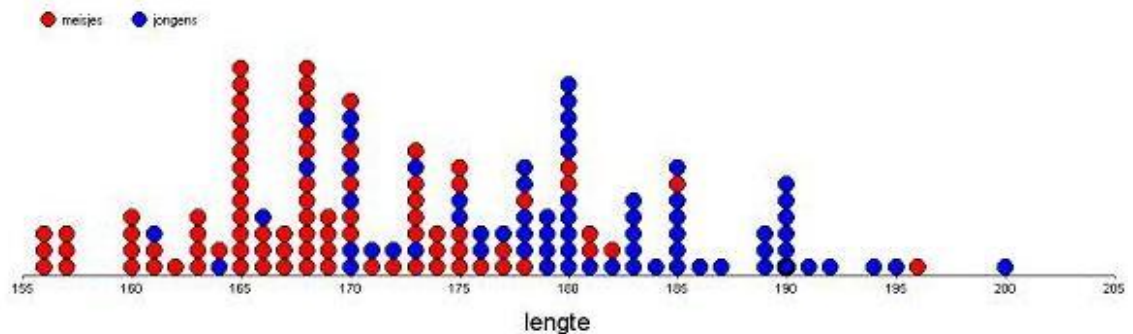
📄 Opgave 9

- 196 cm; 200 cm.
- 165 cm en 168 cm komen beide 13 keer voor.
- Het gemiddelde zegt wel iets over het verschil tussen jongens en meisjes maar niet genoeg. Het gemiddelde wordt erg beïnvloed door uitschieters.



Opgave 10

a.



- b. Zie opgave 9.
- c. De meetgegevens zijn gesorteerd en het geslacht is met een kleur aangegeven.
- d. Mensen ronden graag af naar een makkelijk te onthouden getal.
- e. Voordeel: je kunt soms makkelijker in 1 oogopslag de waarden aflezen.
Nadeel: als steelgetal en bladgetallen niet makkelijk samen te stellen zijn, is een steelbladdiagram slecht te lezen.

Opgave 11

- a. Kwalitatief: lievelingseten; kwantitatief: hoeveelheid zakgeld per maand.
- b. Nee, want er wordt geen hoeveelheid mee aangegeven.
- c. Eenvoudig te meten, nauwkeuriger of onnauwkeuriger heeft weinig zin.
- d. Linker weegschaal: analoog en continu, elke meetwaarde kan worden afgelezen.
Rechterweegschaal: digitaal en discreet, het weergegeven gewicht worden afgerond (bijv. per 10 of 100 gram).
- e. De gemiddelde maandtemperatuur geeft 1 meetwaarde per maand, dus verspringt van maand tot maand en tussenliggende waarden hebben geen betekenis.

Opgave 12

- a. Kwantitatief, discreet, 1900 t/m 2015.
- b. Kwantitatief, continu, -25 t/m (ongeveer) 0.
- c. Kwalitatief, waarden 1 t/m 3.
- d. Kwantitatief, continu, 1 (pasgeboren) t/m 60 gram.
- e. Kwantitatief, discreet (nl. alleen cijfer met 1 decimaal), 1,0-1,1-1,2-1,3-.....-9,9-10,0.
- f. Kwalitatief, CM-EM-NG-NT.
- g. Kwalitatief, 1 t/m 5.



Opgave 13

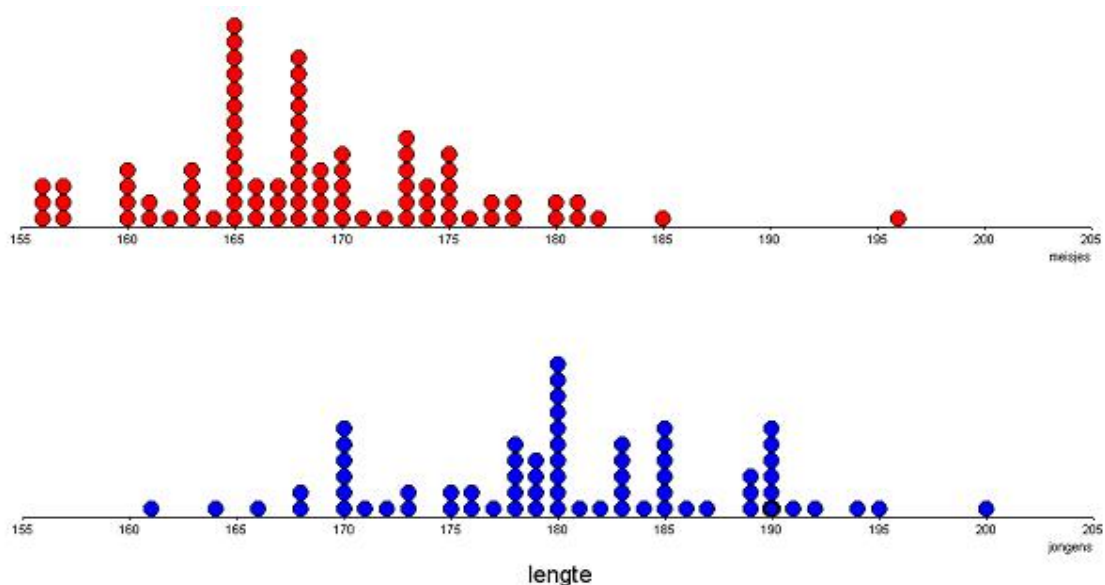
- Ja; volgorde niet van belang; wel tussenruimte.
- Ja; volgorde wel van belang; geen tussenruimte.
- Geboortejaar: niet zo zinvol, omdat er maar vier jaren voorkomen.
- Ja; volgorde wel van belang; wel tussenruimte.
 - Dagelijkse reistijd naar school*: kwantitatief, continu, volgorde belangrijk, afgerond op hele minuten, staafdiagram wel zinvol (geen tussenruimte).
 - Aantal uren per week*: kwantitatief, discreet, volgorde belangrijk, niet afgerond, staafdiagram wel zinvol (tussenruimte mag).
 - Muziekvoorkeur*: kwalitatief, volgorde niet belangrijk, staafdiagram wel zinvol (tussenruimte mag).
 - Belang van bewegingsonderwijs voor iedereen*: kwalitatief, volgorde van belang, staafdiagram wel zinvol (tussenruimte mag).

Opgave 14

- Geslacht*: kwalitatief, m/v (of j/m, of 0=vrouw, 1=man).
- Geboortejaar*: kwantitatief, discreet, 1990 t/m 1994 (of 90 t/m 94).
- Geboortemaand*: kwalitatief, januari t/m december (of 1 t/m 12).
- Gewicht*: kwantitatief, continu, (ongeveer) 60 t/m 100 kg.
- Lengte*: kwantitatief, continu, (ongeveer) 150 t/m 210 cm.
- Cijfergemiddelde*: kwantitatief, continu, 1 t/m 10.
- Cijfer voor wiskunde* (heel cijfer): kwantitatief, discreet, 1 t/m 10.
- Huiswerk* (aantal uren per week): kwantitatief, discreet, 0 t/m 25 (?).
- Wiskundegroep*: kwalitatief, wiskunde A-wiskunde B (eventueel nog \pm wiskunde D).
- Profiel*: kwalitatief, CM-EM-NG-NT (en eventueel combinatieprofielen).
- Plezier*: kwalitatief, 1=nee, 2=weinig, 3=neutraal, 4=behoorlijk, 5=veel.

Opgave 15

- Eigen antwoord.
-



Jongens zijn langer, maar niet allemaal.

Opgave 16

- ICT.
- Omdat de aantallen jongens (69) en meisjes (85) ongelijk zijn.
- Als je een klassenindeling maakt waarbij 180 cm precies een grens is, kun je dit uit je staafdiagram bepalen; bijv.:
jongens: $1,4 + 1,4 + 7,2 + \dots + 1,4 + 1,4 = 42,9 \approx 43$ (%);
meisjes: $2,4 + 1,2 + 1,2 + 1,2 = 6$ (%).
- Meisjes: (van 156 t/m 168 cm, dus) tussen 155 en 169 cm;
jongens: (van 161 t/m 180 cm, dus) tussen 160 en 181 cm.
- Meisjes: (van 173 t/m 196 cm, dus) tussen 172 en 197 cm;
jongens: (van 185 t/m 200 cm, dus) tussen 184 en 201 cm.
- De deelgroep jongens is groter dan de deelgroep meisjes.

Opgave 17

- ICT.
- Eigen antwoord.
- Relatieve frequentietabel/staafdiagram/frequentiepolygoon met klassen van 5 cm.
- Eigen antwoord; in ieder geval een klassenindeling gebruiken waarbij 182 een klassegrens is.
- Beter overzicht van de gehele verdeling.
- Precieze waarden raken verloren.

Opgave 18

- Klassenbreedte 1,0; klassenmiddens 7,0 - 8,0- enz.
(eigenlijk loopt de eerste klasse van 6,45 tot 7,45 dus klassenmiddens 6,95; en dan 7,95; enz.).
- Klassenbreedte 5; klassenmiddens 22,5 - 27,5 - enz.
- Klassenbreedte 50; klassenmiddens 225,5.
- Lengte, gewicht, gemiddelde cijfer, huiswerk.
Dit zijn kwalitatieve variabelen met een redelijke spreiding.

Opgave 19

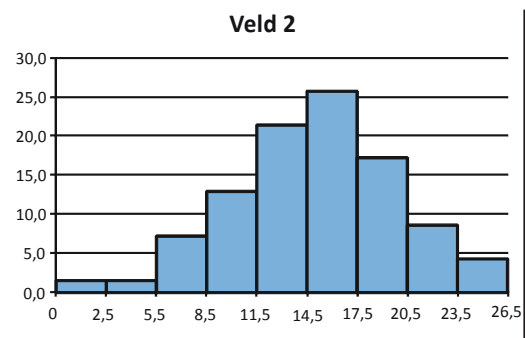
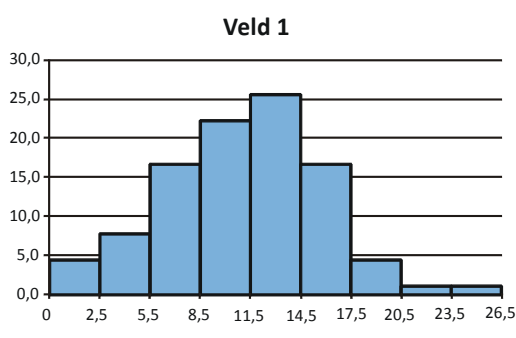
- ICT.
- De twee deelgroepen zijn niet even groot.
- Zie vraag 9 e en f.
- ICT.
- De deelgroep meisjes besteedt meer tijd aan huiswerk dan de jongens.
- Eigen antwoord.



Opgave 20

- a. Kwantitatief, continu.
- b. 26 cm (of eigenlijk 26,4 cm).
- c. Zinvol als in beide velden niet evenveel regenwormen zijn gemeten.
- d.

lengte regenworm (in cm)	aantal (%) veld 1	aantal (%) veld 2
0-2	4,4	1,4
3-5	7,8	1,4
6-8	16,7	7,1
9-11	22,2	12,9
12-14	25,6	21,4
15-17	16,7	25,7
18-20	4,4	17,1
21-23	1,1	8,6
24-26	1,1	4,3



- e. In het eerste veldje iets meer kortere regenwormen; in het tweede veldje iets meer langere regenwormen.

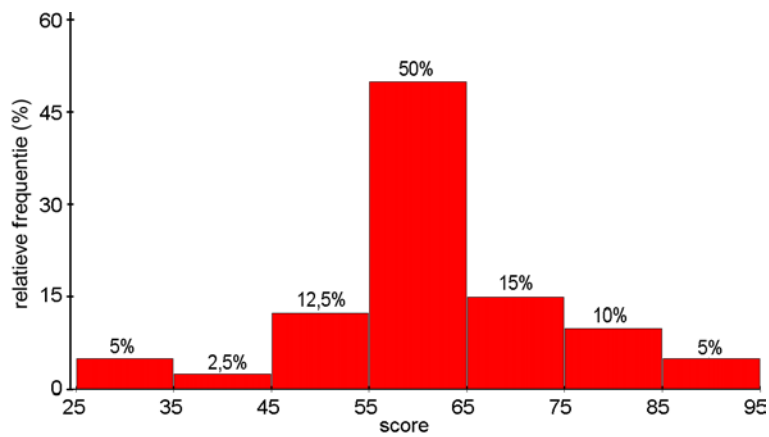


Opgave 21

- a. Kwantitatief, discreet.
b.

Score	Freq.
25 - 34	2
35 - 44	1
45 - 54	5
55 - 64	20
65 - 74	6
75 - 84	4
85 - 94	2
Totaal	40

c.



- d. 80%.
e. De grens van voldoende/onvoldoende valt dan in een klasse en de hoeveelheid onvoldoendes/voldoendes is dan niet meer af te lezen.

Opgave 22

- a. Het betreft de som van alle waarden die kleiner zijn dan de rechter klassegrens.
b. Check je eigen cumulatieve frequentiepolygonen; als het goed is, is aan beide voorwaarden voldaan.



Opgave 23

- ICT.
- Voordeel: je kunt meteen Q1, mediaan en Q3 aflezen. Nadeel: de precieze aantallen gaan verloren.
- 25.
- Wouter behoort tot de 25% zwaarste leerlingen;
Wouter behoort tot de 25% leerlingen met het laagste cijfergemiddelde.
- Aanname: de gewichten zijn evenredig over de klasse verdeeld; rest eigen antwoord.

Opgave 24

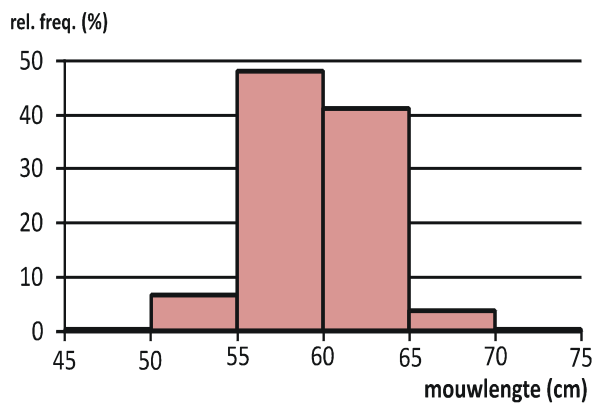
- ICT.
- Eigen antwoord (vergelijk met medeleerlingen).

Opgave 25

- ICT.
- Teken een horizontale lijn op hoogte 50%; de horizontale coördinaat van het snijpunt van deze lijn met de cum. freq. polygoon, is de mediaan.
- Zoals b. Maar dan op hoogte 25% en 75%.
- Zelf doen (gebruik de horizontale as en min, Q1, mediaan, Q3 en max).
- Zie d.

Opgave 26

- Kwantitatief, discreet.
-



- Voordeel staafdiagram: beter overzicht. Voordeel tabel: je hebt alle individuele gegevens.
- $195 / 5001 * 100\% \approx 3,9\%$ (of 4%).
- $89 / 5001 * 100\% \approx 1,8\%$ (of 2%).

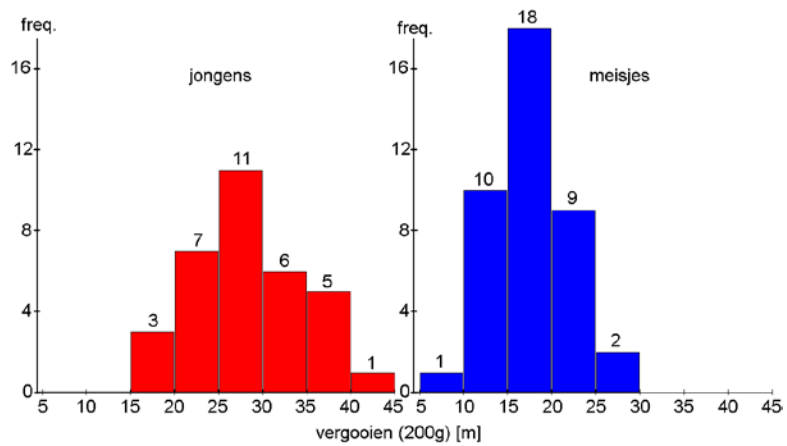
Opgave 27

- 5; 2,5.
- Nee, laatste twee klassen zijn anders.
- Omdat je op die manier bij de verschillende prognoses ook de groei van de totale omvang van de Nederlandse bevolking kunt zien en kunt vergelijken.
- T.o.v. totaal: dan kun je ook zien hoe de verhouding mannen/vrouwen per leeftijdscategorie is afzonderlijk: je kunt voor beide seksen zien hoe de verdeling over de leeftijdscategorieën is.
- De 'top' van de grafiek van de vrouwen is breder dan bij de mannen.
- Categorie 'groen' *Regional Communities*, want dan is het bovenste deel van de grafiek het breedst in verhouding tot het onderste deel.

Opgave 28

- Geslacht*: kwalitatief.
Leeftijd: kwantitatief, discreet.
Sprint: kwantitatief, continu.
Verspringen: kwantitatief, continu.
Vergooien: kwantitatief, continu.
- (Om te vergelijken kun je eigenlijk beter relatieve frequenties gebruiken, maar omdat het verschil hier zo duidelijk is maakt het hier niet echt uit.)
-

Vergooie n 200g [m]	Jongen		Meisje	
	Fre q.	%	Fre q.	%
5,0-9,9	0	0,0	1	2,5
10,0- 14,9	0	0,0	10	25, 0
15,0- 19,9	3	9,1	18	45, 0
20,0- 24,9	7	21, 2	9	22, 5
25,0- 29,9	11	33, 3	2	5,0
30,0- 34,9	6	18, 2	0	0,0
35,0- 39,9	5	15, 2	0	0,0
40,0- 44,9	1	3,0	0	0,0
Totaal	33	10 0	40	10 0



- d. De jongens gooien over het algemeen verder dan de meisjes.



§ 2.2 Verbanden tussen datarepresentaties

Opgave 29

- Zelf doen.
- Zelf doen.
- Ze hebben allemaal dezelfde vorm.

Opgave 30

- Ja, met moeite; nee, niet direct (beide antwoorden zijn correct).
- Nee, je weet niet hoe de lengtes per klasse precies verdeeld zijn.

Opgave 31

- Korter, want er zijn links veel meer bolletjes bij elkaar dan rechts.
- Korter, zie a.
- Ja, het bolletje dat helemaal rechts zit, zou een uitschieter kunnen zijn.

Opgave 32

De derde: de meeste bolletjes zitten links, dus aan de linkerkant moeten de sommen telkens flink stijgen, terwijl de sommen aan de rechter kant heel weinig zullen stijgen (weinig bolletjes).

Opgave 33

In de frequentietabel mist het specifieke turfje of bolletje voor iedere waarneming apart, dus bij een dotplot heb je meer informatie.

Opgave 34

- Dezelfde (en dus evenveel) klassen; beide geven de frequenties per klasse; beide hebben géén speciaal nummertje/tekentje voor iedere waarneming afzonderlijk.
- Beschrijving schetsje aan de hand van de eerste klasse: in plaats van één klasse (8,0 - 8,4) krijg je vijf klassen tussen 8,0 en 8,5 en ieder van de vijf klassen heeft dezelfde frequentie, nl $10 / 5 = 2$. Je verdeelt de tien waarnemingen tussen 8,0 en 8,5 dus evenredig over de vijf klassen: je weet niet hoe de waarnemingen in werkelijkheid waren verdeeld over deze vijf klassen (als je alleen dit staafdiagram met klassen van 0,5 seconde hebt) en daarom ziet dit schetsje er anders uit dan het staafdiagram in opgave 1.

Opgave 35

In de diagrammen met klassenindelingen kun je makkelijker direct zien wat de meest gesprinte tijden waren. In de dotplot en de frequentietabel uit opgave 5 moet je daar meer moeite voor doen. De klassenindeling werkt als een samenvatting en dat maakt het makkelijker om dit snel af te lezen.

Opgave 36

- De 25% langzaamste leerlingen liepen de sprint tussen de 9,5 en de 11,3 seconden.
- De allerlangzaamste leerling liep hem dus in 11,3 seconden.
- De allerlangzaamste leerling kun je ook direct uit de dotplot aflezen; voor de 25% langzaamste leerlingen moet je heel veel meer moeite doen, terwijl de boxplot dat op een presenteerblaadje aanreikt.

Opgave 37

a.

Representatie	Esmee's tijd is direct af te lezen	
	Ja/nee	Zo ja: hoe? Zo nee: waarom niet?
Dotplot	Ja	Esmee heeft haar eigen 'dot'.
Frequentietabel	Beide	Je weet wel bij welk groepje Esmee hoort, maar je kunt niet apart haar eigen meting zien.
Staafdiagram (histogram)	Beide	Je weet wel in welk staaf Esmee's sprinttijd staat, maar je kunt niet apart haar eigen meting zien.
Frequentiepolygoon	Beide	Je weet wel bij welke punt Esmee's sprinttijd hoort, maar je kunt niet apart haar eigen meting zien.
Bovenstaande diagrammen met klassenindeling	Beide	Bij de dotplot blijft Esmee haar eigen 'dot' houden. Maar voor de anderen geldt dat je zelfs Esmee's specifiek sprinttijd niet meer ziet, want deze valt in een klasse van tijden.
Steelbladdiagram	Ja	Esmee heeft haar eigen blaadje.
Cumulatief frequentiepolygoon	Beide	Je weet wel bij welke punt Esmee's sprinttijd hoort, maar je kunt niet apart haar eigen meting zien.
Boxplot	Nee	Je kunt alleen intervallen van sprinttijden aflezen.

b.

Representatie	Direct afleesbaar hoeveel klasgenoten sneller zijn dan Esmee	
	Ja/nee	Zo ja: hoe? Zo nee: waarom niet?
Dotplot	Nee	Je moet de 'dotjes' eerst tellen.
Frequentietabel	Nee	Je moet eerst frequenties optellen.
Staafdiagram (histogram)	Nee	Je moet eerst frequenties optellen.
Frequentiepolygoon	Nee	Je moet eerst frequenties optellen.
Bovenstaande diagrammen met klassenindeling	Nee	Je moet eerst frequenties optellen en je weet niet hoeveel tijden in de klasse van Esmee kleiner zijn dan haar tijd.
Steelbladdiagram	Nee	Je moet eerst blaadjes optellen.
Cumulatief frequentiepolygoon	Ja	De waarde op de verticale as geeft het aantal dat sneller is.
Boxplot	Ja	Althans: in percentagedelen van 25%.

Opgave 38

Nee, dat is niet mogelijk. Neem bijv. de 26 waarnemingen in klasse 9,5-<10 seconden. Je hebt geen idee welke waarde deze 26 waarnemingen precies hebben, anders dan dat ze allemaal tussen de 9,5 en 10 seconden liggen. Het kan best zijn dat er 25 keer 9,5 seconden en 1 keer 9,9 seconden is gesprint; het kan ook zijn dat er 26 keer 9,8 seconden is gesprint etc.

Anders gezegd: je kunt de waarde van de blaadjes voor het steelbladdiagram niet aflezen uit dit staafdiagram.

Opgave 39

- ICT.
- In een frequentietabel kun je direct per steelgetal de frequentie voor dat steelgetal aflezen (in het steelbladdiagram moet je zelf tellen). Als er heel veel blaadjes per steelgetal zijn, is een frequentietabel overzichtelijker. Als de steelgetallen niet zinvol verder kunnen worden opgedeeld, geeft de frequentie per steelgetal evenveel informatie als het steelbladdiagram.
- ICT.
- In een frequentietabel kun je direct de frequentie aflezen; in het staafdiagram zie je in één oogopslag de modus of juist de minst getelde waarneming.

Opgave 40

- ICT.
- In het cumulatieve frequentiepolygoon kun je ook direct de somfrequenties aflezen voor andere waarden dan de 25%, 50% en 75%: je ziet bijv. dat er tien mensen in minder dan 8,5 seconden gesprint hebben. In de boxplot kun je direct de interkwartielafstand aflezen en of deze meer links, meer in het midden of juist meer rechts ligt, omdat Q1 en Q3 direct zichtbaar zijn.

Opgave 41

- Je weet de werkelijke getallen niet; bijv. de eerste twee werknemers kunnen beide 400, maar ook beide 450 euro verdienen, of beide een ander verschillend loon.
- Klassenmiddens: $425 - 475 - 525 - 775$ $\frac{425 \cdot 2 + 475 \cdot 3 + \dots + 775 \cdot 1}{2+3+\dots+1} = \frac{14425}{25} = 577$ (€).
- Je weet bijv. alleen dat de mediaan zit in de klasse 550-<600, maar de mediaan kan dus elke waarde in dit interval zijn; evenzo voor het minimum, maximum en de kwartielen.

Opgave 42

Met de *dotplot* komt Tjeerds sprong niet echt tot z'n recht en dit geldt ook voor *steelbladdiagram* en *frequentietabel zonder klassenindeling* (ofwel met klassen van 1 cm breed); ook *frequentiepolygoon* en *staafdiagram zonder klassenindeling* (ofwel klassen van 1 cm breed) geven niet direct inzicht in Tjeerds best hoge sprong.

Tjeerd zal blijer zijn met een *staafdiagram* of *frequentiepolygoon met klassenindeling van 30 cm* breed: je kunt dan direct zien dat er veel meer klasgenoten lager sprongen dan Tjeerd en veel minder die hoger sprongen dan Tjeerd. Met de *boxplot* zal hij ook blij zijn: zijn sprong is nl. zo'n beetje gelijk aan Q3 en dus kun je zien dat er maar liefst 75% minder hoog sprongen dan Tjeerd. In een *absolute cumulatieve frequentiepolygoon* kun je ook goed zien hoe hoog Tjeerds sprong is ten opzichte van zijn klasgenoten.

Opgave 43

- ICT.
- Zelf doen.
- ICT.

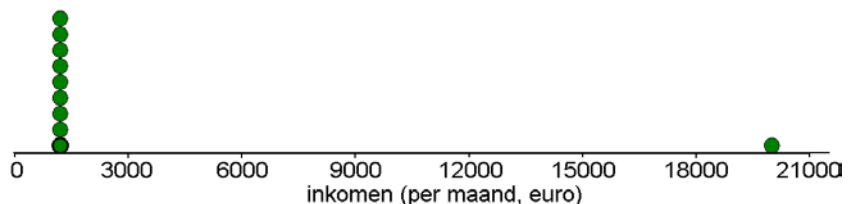
§ 2.3 Frequentieverdelingen typeren

Opgave 44

- Eigen antwoord.
- Meisjes: 168,8 cm; jongens: 180,4 cm.
- Meisjes: 165 cm; jongens: 180 cm.
- Meisjes: 168 cm; jongens: 180 cm.
- Meisjes: modus, mediaan, gemiddelde; jongens: modus/mediaan, gemiddelde.
Bij beiden is het gemiddelde de grootste waarde; bij de meisjes is de modus een stuk kleiner dan de mediaan en het gemiddelde terwijl de modus van de jongens gelijk is aan de mediaan. Bij de meisjes is het verschil tussen de kleinste en de grootste centrummaat veel groter (3,4 cm) dan bij de jongens (0,4 cm).
Bij de meisjes is het dus belangrijk om naar alledrie de centrummaten te kijken!

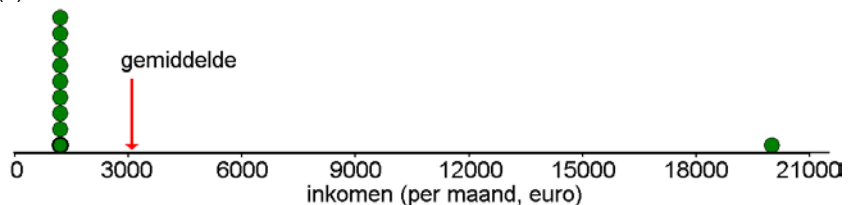
Opgave 45

a.



b. Eigen antwoord.

c. Gemiddelde = 3080 (€).



- Eigen antwoord.
- Zie het als een wipwap met daarop de bolletjes als gewichtjes: als de grafiek bij het gemiddelde wordt ondersteund blijft het in evenwicht.
- Mediaan = 1200.
- 1200 €
- Modus en mediaan wel, gemiddelde is heel anders.

Opgave 46

- Jongens: 39 cm; meisjes: 40 cm.
- Nee: bij weglating van de uitschieters (uitzonderlijke waarnemingen) naar boven is de spreidingsbreedte bij de meisjes wel veel kleiner!

Opgave 47

- Van 161 t/m 176 cm.
- 85 meisjes, dus 25% is 21 (of 22) meisjes; in boxplot vanaf langste meisje terugtellen geeft lengte 173 cm voor 21^e en 22^e leerling; dus Q_3 is 173 cm; het maximum is 196 cm; klopt.
- Jongens: 161 – 176 – 180 – 185 – 200; meisjes: 156 – 165 – 168 – 173 – 126.
- Jongens: 9 cm; meisjes: 8 cm. Nee, slechts 1 cm verschil.

Opgave 48

- a. 2, rechts-scheef.
- b. 1.
- c. 3.
- d. 2.

Opgave 49

- a. Meer symmetrisch (klokvormig) dan uniform of scheef verdeeld; tweekoppig.
- b. Eentoppig; niet uniform, niet duidelijk scheef verdeeld, niet symmetrisch.
- c. Eigen antwoord.

Opgave 50

- a.
 - 1. gem = 179,5 cm; mediaan = 177,5 cm; modus = klassemidden van klasse 175-<180 cm en dat is 177,5 cm. Omdat dit een symmetrische verdeling is, is de modus hier de beste centrummaat.
 - 2. gem = 190 cm; mediaan = 192,5 cm; modus = klassemidden van klasse 195-<200 cm en dat is 197,5 cm. Omdat dit een scheve verdeling is, geeft de mediaan hier de beste indruk.
 - 3. gem = 178,2 cm; mediaan = 177,5 cm; modus = klassemidden van klasse 175-<180 cm en dat is 177,5 cm. Ook hier geeft de mediaan de beste indruk: het is een (bijna) uniforme verdeling.
- b.
 - 1. Spreidingsbreedte = 50 cm; interkwartielafstand = 15 cm;
 - 2. Spreidingsbreedte = 45 cm; interkwartielafstand = 12,5 cm;
 - 3. Spreidingsbreedte = 50 cm; interkwartielafstand = 20 cm.Bij alledrie zegt de interkwartielafstand meer over de verdeling dan de spreidingsbreedte.
- c. Gemiddelde = 191,5 cm; mediaan en modus blijven gelijk; ook nu blijft mediaan beste centrummaat voor de gehele verdeling.
- d. Gemiddelde = 187,2 cm; mediaan blijft wederom gelijk; modus = klassemidden van klasse 170-<175 cm en dat is 172,5 cm; ook nu blijft mediaan beste centrummaat voor de gehele verdeling.

Opgave 51

- a. Gemiddelde: die wordt nu gelijk aan mediaan en modus.
- b. Nee: Q1 en Q3 waren allebei gelijk aan de mediaan en blijven dat ook.
- c. Ja.
- d. Het is slechts een van alle waarnemingen en deze ene waarneming wijkt in z'n eentje enorm af van de andere: het lijkt de uitzondering die de regel bevestigt.



Opgave 52

- a. Bij de meisjes vatten de mediaan en de interkwartielafstand de verdeling het beste samen (gemiddelde en spreidingsbreedte worden beïnvloed door uitschieters) en bij de jongens kun je het beste de combinatie van gemiddelde en mediaan kiezen.
- b. $Q1 = 176$ en kwartielafstand = 9; $176 - 1,5 \cdot 9 = 162,5$; 161 is kleiner.
 $Q3 = 185$ en kwartielafstand = 9; $185 + 1,5 \cdot 9 = 198,5$; 200 is groter.
Dus het zijn beide uitschieters.
- c. Voor de jongens geldt:
centrummaten: mediaan = 180; gemiddelde = 180,4; modus = 180 (cm);
spreidingsmaten: spreidingsbreedte = $195 - 164 = 31$ (cm);
kwartielafstand = $185 - 176 = 9$ (cm).
Meisjes: zelf doen.
- d. Zie vraag a, spreidingsbreedte.
- e. Zie vraag a, gemiddelde.
- f. Als twee verdelingen vrijwel gelijk zijn, op uitschieters na, dan heeft het zin de uitschieters te verwerken in de samenvattingen; in elk ander geval is het beter de uitschieters weg te laten.

Opgave 53

- a. Mediaan = 65 (kg); gemiddelde = 65,2 (kg).
- b. Meerdere gewichten komen 'het vaakst' voor, nl. 65 en 70 kg.
- c. Spreidingsbreedte = $90 - 49 = 41$ (kg);
kwartielafstand = $70,5 - 58,5 = 12$ (kg).
- d. Centrummaat: gemiddelde (van 65,2 naar 64,8);
spreidingsmaat: spreidingsbreedte (van 41 naar 32).
- e. Nee.
- f. Van 49 t/m 58 kg.
- g. 7 van de 69 jongens; dat is $7 / 69 \cdot 100\% \approx 10\%$.



Opgave 54

a. Je weet de ruwe data niet meer;

meisjes: in klasse 55-<60;

jongens: in klasse 65-<70.

b.

Gewicht (kg)	Klassen-midden	Jongens Freq.	Meisjes Freq.
40-44	42	0	3
45-49	47	1	5
50-54	52	7	20
55-59	57	13	30
60-64	62	10	16
65-69	67	14	8
70-74	72	10	0
75-79	77	6	2
80-84	82	6	0
85-89	87	0	0
90-94	92	1	0
95-99	97	0	0
Totaal		68	84

c. Je weet niet hoe de werkelijke waarnemingen over de klassen zijn verdeeld;

d. Jongens: gemiddelde $\approx 65,8$ (kg);

meisjes: gemiddelde $\approx 57,1$ (kg).

e. Jongens: 65,8 versus 65,2;

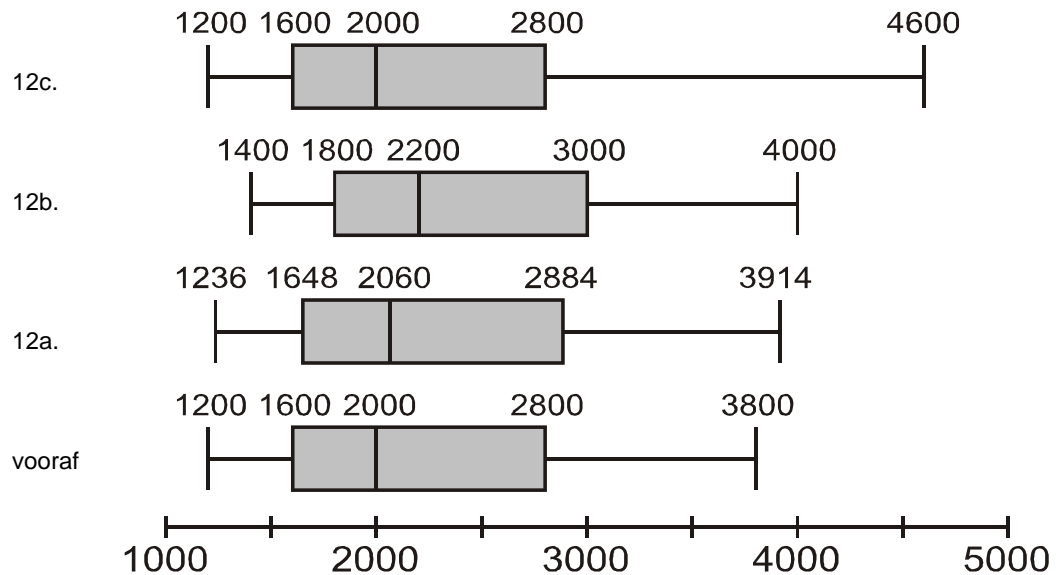
meisjes: 57,1 versus 56,8;

ja, ze wijken enigszins af.



Opgave 55

- Modaal: 1648 euro; gemiddelde: 1854 euro; boxplot zie hieronder.
- Modaal: 1800 euro; gemiddelde: 2000 euro; boxplot zie hieronder.
- Modaal: 1600 euro; gemiddelde: 1807 euro; boxplot zie hieronder.



Opgave 56

- Er zijn meerdere polsslagen die 'het vaakst' voorkomen, dus een echte modus is er niet; niet zinvol.
- Voor: 65,1; na: 75,7; ja.
- Voor: Q1 = 59; mediaan = 64; Q3 = 72;
na: Q1 = 68; mediaan = 74; Q3 = 84;
ja.
- Ja, aan gemiddelde, mediaan en kwartielen zie je dat deze behoorlijk afwijkend is.

Opgave 57

- Dotplot 1: de dot van 25 verplaatsen naar 20.
Dotplot 2: de boxplot is al goed, dus meerdere mogelijkheden (bijv. de dot van 2 verplaatsen naar 1, 3 of 4).
Dotplot 3: de dot van 16 verplaatsen naar 15.
- Zelf doen, bijv. tussen 0 en 20 dotjes.
- Symmetrische en (redelijk) uniforme verdeling.

Opgave 58

- CM: tweetoppig; EM: een beetje symmetrisch (klokvormig);
NG: vrijwel symmetrisch (klokvormig); NT: rechts-scheef.
- Zelf doen.
- CM: gemiddelde $\approx 6,5$; mediaan = 6
EM: gemiddelde $\approx 7,3$; mediaan = 7
NG: gemiddelde $\approx 7,5$; mediaan = 7
NT: gemiddelde $\approx 8,1$; mediaan = 8.
- Bij NT, ja (geen staart, geen uitschieters).

Opgave 59

- IQ:* redelijk symmetrisch, nogal wat uitschieters, ...
Gem temp: scheef met staart links, ...
dhagef: redelijk symmetrisch met uitschieters rechts, beetje scheef, ...
Eruptieduur: tweetoppig, ...
d-10: erg scheef, met staart rechts, ook uitschieters rechts, ...

Opgave 60

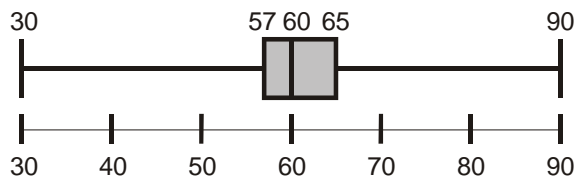
- Alle huishoudens* (lichtblauw): scheef met staart rechts;
Paar met kinderen: scheef met staart rechts;
Paar zonder kinderen: nog veel schever met lange staart rechts;
Alleenstaande: symmetrisch met één top (klokvorm).
- Lastig aflezen... spreidingsbreedte *Alle huishoudens:* $\pm 100\ 000$ euro;
Paar met kinderen: $\pm 90\ 000$ euro;
Paar zonder kinderen: $\pm 70\ 000$ euro;
Alleenstaande: $\pm 50\ 000$ euro.
- Alle huishoudens:* van klein naar groot: modaal–mediaan–gemiddelde.
Paar met kinderen: van klein naar groot: modaal–mediaan–gemiddelde.
Paar zonder kinderen: van klein naar groot: modaal–mediaan–gemiddelde (verder uit elkaar dan bij *Paar met kinderen*, want schevere verdeling).
Alleenstaande: modaal, mediaan en gemiddelde ongeveer gelijk.
- Van *Alleenstaanden* is t.o.v. *Paar zonder kinderen* de spreidingsbreedte kleiner en de modus lager. Voor *Paar zonder kinderen* t.o.v. *Paar met kinderen* geldt hetzelfde. De modus van *Alleenstaanden* komt veel vaker voor dan de modus van *Paar zonder kinderen*, en die komt iets vaker voor dan de modus van *Paar met kinderen*.

Opgave 61

- De variabele *profiel* is niet kwantitatief.
- Er is geen ordening tussen de profielen, dus er is ook geen 'middelste'.
- Eigen antwoord.
- Het ligt eraan hoe je de profielen ordent.

Opgave 62

- 61,0.
-



- Mediaan, want er zijn nogal wat uitschieters. (Maar omdat er 'toevallig' ongeveer evenveel uitschieters naar boven als naar onderen zijn, is hier het gemiddelde ook redelijk geschikt.)
- De werkelijke metingen worden bij de berekening vervangen door de klassenmiddens. En omdat de klassen breder zijn kan elke meting meer afwijken van het midden van de klasse waarin het ligt.

Opgave 63

- a. Maandag; uitschieters hebben een te grote invloed.
- b. Zaterdag.
- c. Ongeveer 50% (net iets minder).
- d. Waarschijnlijk op de vrijdag, maar zeker weet je het niet
(hoogst mogelijk aantal geboorten op maandag (ongeveer): $13 * 415 + 13 * 425 + 13 * 450 + 13 * 520 = 23.530$; op vrijdag (ongeveer): $13 * 420 + 13 * 430 + 13 * 450 + 13 * 470 = 23.000$; dus theoretisch kan het op maandag net iets meer zijn, maar dit is zeer onwaarschijnlijk).
- e. De modus kan bijv. 425 zijn; als 425 geboorten op elke dag bijvoorbeeld 2 keer voorkomt en elk ander aantal slechts 1 keer.
- f. Nee, het gemiddelde op vrijdag is zeker groter dan het gemiddelde op zondag;
vrijdag: gemiddelde groter dan $(13 * 420 + 13 * 430 + 13 * 450 + 13 * 470) / 52 \approx 440$;
zondag: gemiddelde kleiner dan $(13 * 385 + 13 * 405 + 13 * 420 + 13 * 440) / 52 \approx 412$.

Opgave 64

- a. Gemiddelde $\approx 21,82$; mediaan = 20,0; modus = 16,0;
spreidingsbreedte = $40,0 - 5,0 = 35,0$; kwartielafst. = $26,50 - 16,0 = 10,50$.
- b. Je hebt dan de precieze gegevens en metingen worden niet benaderd door een klassenmidden, dus is nauwkeuriger.
- c. ...



§ 2.4 Twee groepen vergelijken

Opgave 65

- De frequentieverdeling van het NT-profiel is klokvormiger dan die van het CM-profiel. Het CM-profiel is uniformer dan het NT-profiel.
De modus van het NT-profiel ligt meer naar rechts dan de modus van het CM-profiel: deze ligt erg links.
De verdeling van het CM-profiel loopt langer door naar rechts dan in het NT-profiel.
- In het NT-profiel worden de meeste cijfers rondom de modus gescoord: weinig echt lage cijfers en weinig echt hoge cijfers
In het CM-profiel komen alle cijfers ongeveer even vaak voor; er zijn echter extra veel lage cijfers (de modus) en de allerhoogste cijfers (het meest rechts) van beide groepen samen worden in het CM-profiel gescoord.

Opgave 66

- NT-profiel: gemiddelde is 6,6 en mediaan is 6,6.
CM-profiel: gemiddelde is 6,5 en mediaan is 6,7.
- NT-profiel is symmetrisch verdeeld; CM-profiel is iets meer links-scheef verdeeld (in verhouding meer lage cijfers dan in het NT-profiel). Binnen de 50% hoogste scores in het NT-profiel zitten lagere cijfers (tussen 6,6 en 6,7) dan binnen de 50% hoogste scores in het CM-profiel.

Opgave 67

- De box van de boxplot van het CM-profiel is veel langer dan de box van het NT-profiel, vooral aan de linkerkant; aan de linkerkant beginnen de boxplots tegelijk, de boxplot van het CM-profiel loopt langer door naar rechts; de medianen liggen ongeveer gelijk.
- Zowel de spreidingsbreedte als de interkwartielafstand van het NT-profiel zijn kleiner dan die van het CM-profiel. De laagste 50% begint en eindigt bij beiden (ongeveer) bij dezelfde cijfers maar in het CM-profiel is de kleinste 25% veel minder verspreid dan de 25% daarboven, terwijl dat bij het NT-profiel net andersom is: in het CM-profiel worden meer echt lage cijfers gescoord. Al eerder is opgemerkt dat de echt hoge cijfers ook in het CM-profiel worden gescoord.

Opgave 68

- Het CM-profiel is gelijkmatiger (uniformer) dan het NT-profiel, op een sprong tussen de 5 en de 5,5 na. Het NT-profiel heeft opeens een sprong tussen de 6,5 en 7. En is eerder bij de 100%, al bij de 8,5 (CM-profiel bij de 9).
Je kunt ook in één oogopslag zien dat ongeveer 75% een 7 of lager scoort bij het NT-profiel terwijl bij het CM-profiel ruim 60% een 7 of lager scoort...
- Zie antwoorden bij opgave 1 t/m 3.
- Eigen antwoord.

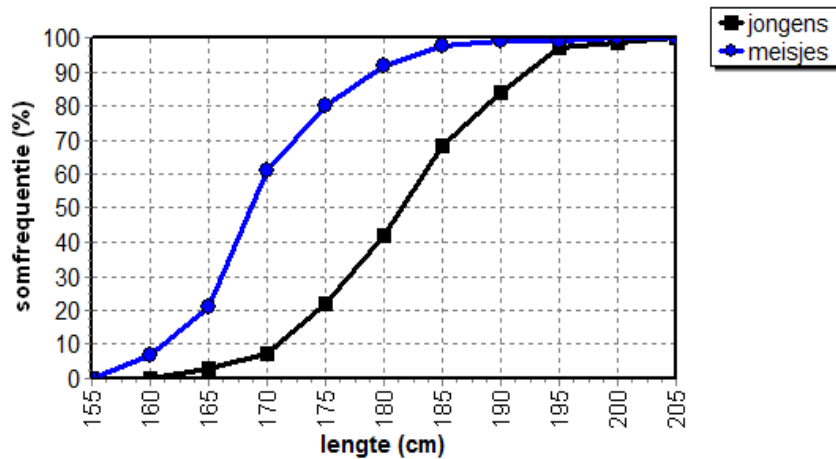
Opgave 69

- Beide polygonen springen niet opeens links of rechts met een sprong omhoog.
- ICT of terugzoeken in vorige opgaven.
- Dat kan correct zijn als hierbij gebruik gemaakt is van de klassen 4-<5 etc.
- Ruim 40%.



Opgave 70

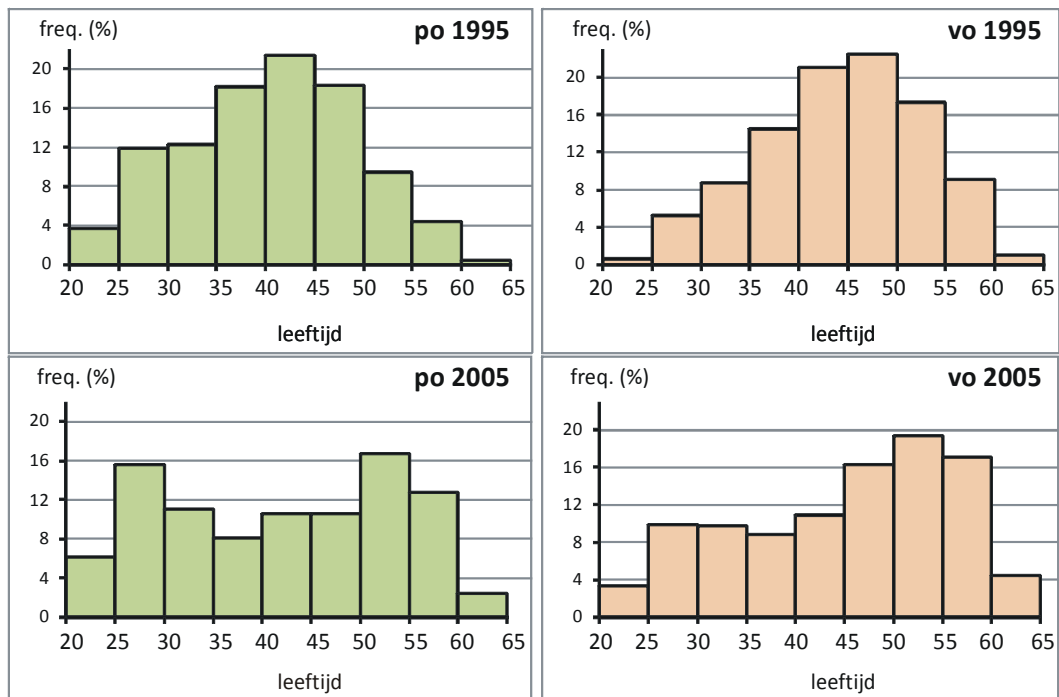
a. ICT, bijv.



b. Mediaan bij 50%; Q1 bij 25% en Q3 bij 75%.

Opgave 71

a. en d. ICT, bijv.



b. De verdeling van 1995 heeft één top en die van 2005 heeft twee toppen. De linker top betreft de nieuwe instromers in het onderwijs.

c. Klassenmiddens: 22,5-27,5-...- 62,5.

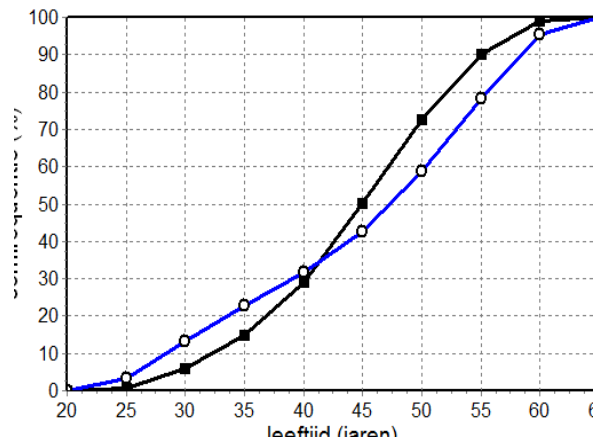
Po 1995: gemiddelde $\approx 40,5$; po 2005: gemiddelde $\approx 42,3$.

e. De verdeling in 1995 is redelijk klokvormig terwijl die van 2005 gelijkmatiger/uniformer wordt.

De staven tussen 25 en 45 zijn ongeveer even hoog: er komt in die leeftijdsklassen niemand bij.

De staven tussen 45 en 60 zijn lager: er zijn minder mensen werkzaam in die leeftijdsklassen dan in 1995. In totaliteit komt is er dus sprake van uitstroom.

f.



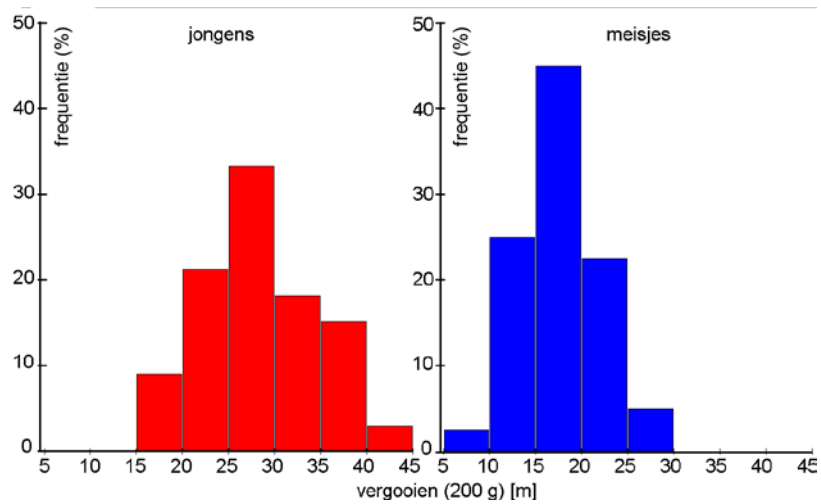
Teken de boxplots er zelf bij!

- g. Het uniformer worden van de verdeling zie je terug in de rechttere polygoon van 2005. Aan de medianen (bij 50%) zie je dat er minder jongeren zijn in 2005 (50% jonger dan 47/48 jaar) dan in 1995 (50% jonger dan 45 jaar). De interkwartielafstand is in 2005 groter geworden dan in 1995: de jongste 25% is in 2005 wat jonger is dan in 1995 en de oudste 25% is wat ouder dan in 1995.



Opgave 72

- a. Jongens: gemiddelde $\approx 27,6$; mediaan = 27,0; modus = 25,0 (m);
 jongens: spreidingsbreedte = 24; kwartielafstand = 7,5 (m);
 meisjes: gemiddelde $\approx 17,1$; mediaan = 16,5; modus = 16,0 (m);
 meisjes: spreidingsbreedte = 22,5; kwartielafstand = 5,75 (m).



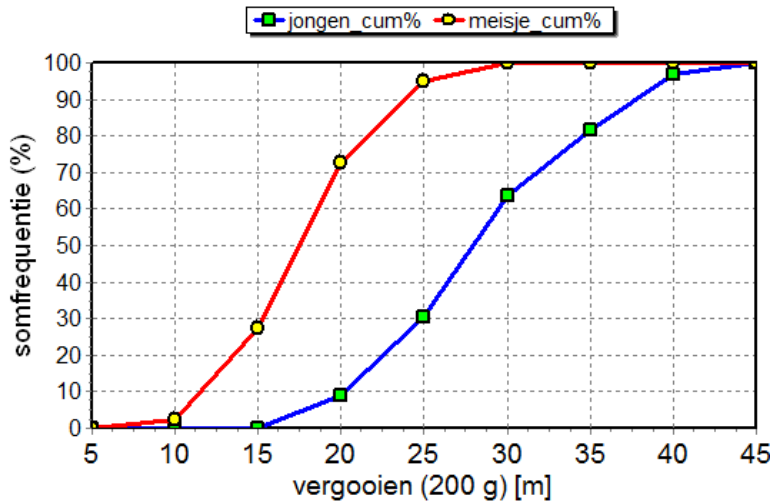
Rest zelf.

Conclusies: de groep jongens gooit verder dan de groep meisjes

(zij gooien tussen de 15 en 45 meter, terwijl de meisjes tussen de 5 en 30 meter gooien).

Bij de meisjes zijn de afstanden symmetrischer en minder verspreid verdeeld dan bij de jongens: in de groep jongens heb je een staart bij en wat meer links-scheve verdeling.

b.

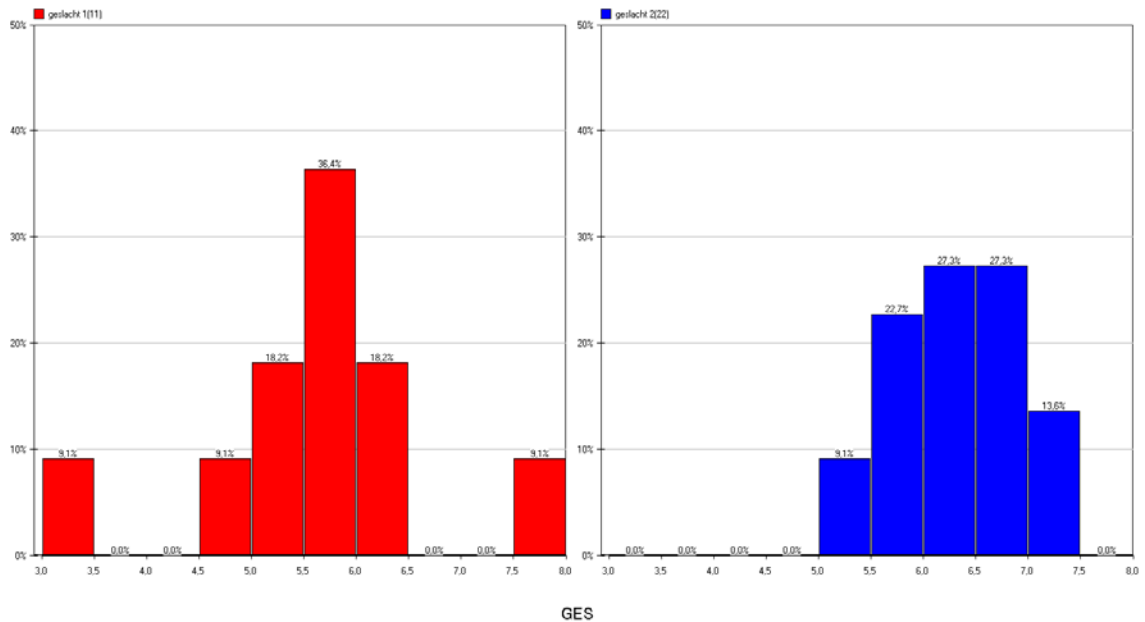


Nu zie je in één oogopslag dat de groep jongens verder gooit dan de groep meisjes.



Opgave 73

a. ICT, bijv.

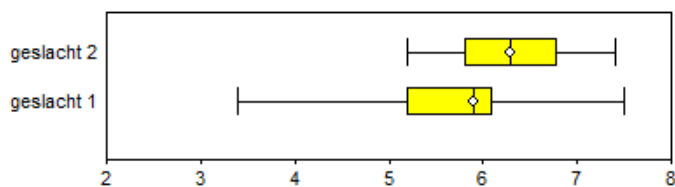


Er vallen drie dingen op:

- *Geslacht 1* heeft een grotere spreidingsbreedte.
- *Geslacht 1* heeft een hoge modus en *geslacht 2* heeft geen modus.
- Bij *geslacht 2* scoort niemand lager dan een 5 (maar ook niet hoger dan 7,5).

b.

Variabele: GES



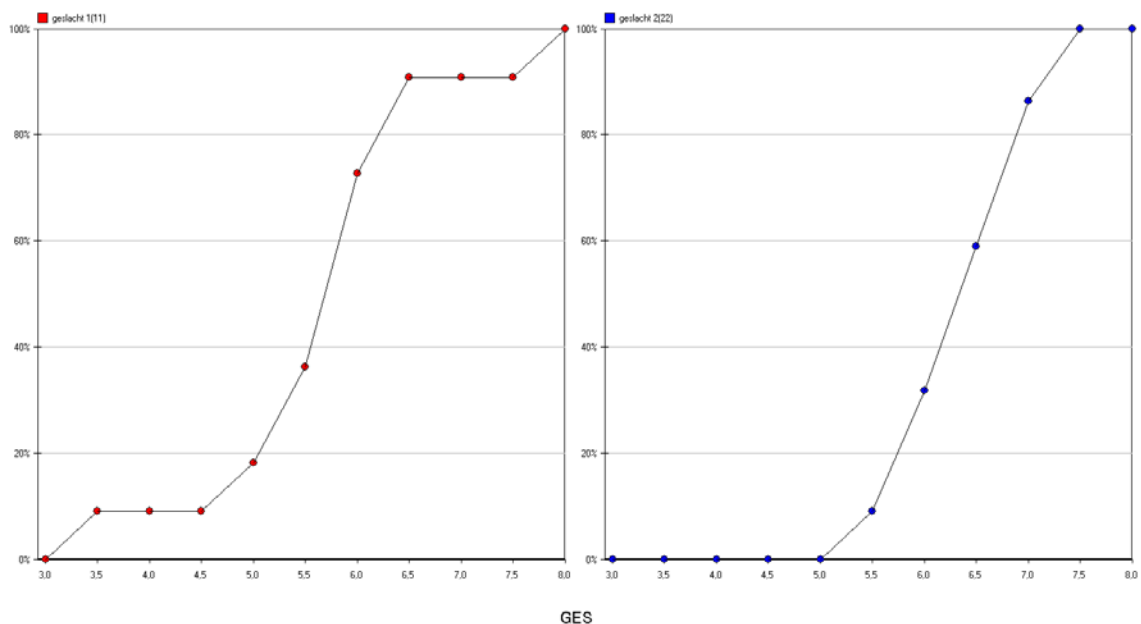
Ook hier zie je het verschil in spreidingsbreedte terug en de ligging van de cijfers.

Verder valt op:

- De mediaan van *geslacht 1* is lager dan de mediaan van *geslacht 2*.
- Maar liefst 25% van *geslacht 1* heeft lager gescoord dan alle betrokkenen van *geslacht 2*.



C.



Eigen antwoord; vergeet niet dat het hier alleen om de jongens en meisjes in deze ene klas gaat: het betreft zeker geen representatieve steekproef van alle jongens en meisjes!

Opgave 74

- De boxplot van type A ligt in zijn geheel rechts van de drie andere boxplots (de kleinste waarde van type A is groter dan de grootste waarden van type B, C en D).
 - Het zullen steekproeven zijn waarop deze boxplots zijn gebaseerd en natuurlijk nooit de voltallige populatie van lampen ooit gemaakt of nog te maken! En een steekproef kan door toeval afwijken van de totale populatie.
 - Het aantal branduren van de lamp van type B met de kortste brandtijd is even groot als het derde kwartiel van type C: 25% van type C brandt langer dan de lamp van type B met de kortste brandtijd.
 - Wat je veilig kunt zeggen:
 - Het kortste aantal branduren is bij beiden gelijk: of je nu een lamp van type C of van type D hebt, ze branden minimaal 2000 uur.
 - De mediaan van type C is gelijk aan het derde kwartiel van type D: de kortste 75% van type D gaat even kort mee als de 50% kortste van type C.
 - Het eerste kwartiel van type C valt samen met de mediaan van type D: de kortste 50% van type D gaat even kort mee als de kortste 25% van type C.
- => Al met al lijkt je met type C iets meer kans te hebben op meer branduren dan met type D.
In het algemeen – zeker als de steekproeven niet zo heel groot zijn – kun je bij dit soort geheel overlappende boxplots nauwelijks met zekerheid uitspraken doen.

Opgave 75 en 76

...

§ 2.5 Samenhang tussen twee variabelen

Opgave 77

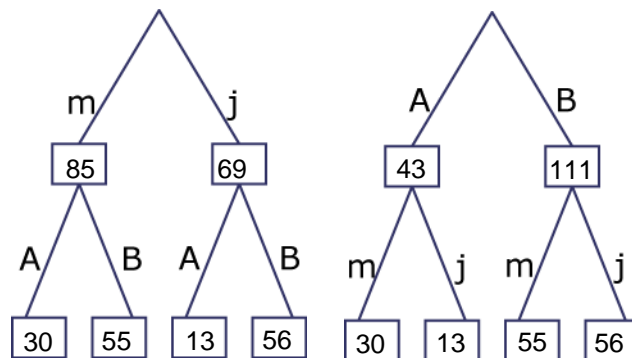
- Grofweg lijkt het volgende verband te bestaan: als de leeftijd van de mannen stijgt, dan stijgt dat van de vrouwen ook. Je ziet dit doordat de puntenwolk grofweg een stijgende rechte lijn vormt.
- Tussen de 75 en 80 jaar (ietsje hoger dan 73).

Opgave 78

- ICT, ...
- Eigen antwoord.

Opgave 79

- 55 meisjes.
-

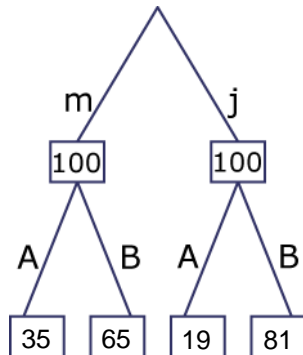


- $(55 / 85) * 100\% \approx 64,7\%$.
- $56; (56 / 69) * 100\% \approx 81,2\%$.
- $(56 / 111) * 100\% \approx 50,5\%$; het rechter diagram.



Opgave 80

a.



b.

Geslacht	Wiskundegroep		Totaal
	A	B	
J	18,84 %	81,16 %	100 %
M	35,29 %	64,71 %	100 %
Totaal	27,92 %	72,08 %	100 %

c. Ruim 29%; ...

d. Bij de jongens verschillen de percentages nog meer.

Opgave 81

a. $(55 / 111) * 100\% \approx 49,5\%$.

b. Je splitst eerst in wiskunde A en wiskunde B en kijkt daarna naar het geslacht.

c.

Geslacht	Wiskundegroep		Totaal
	A	B	
J	30,23 %	50,45 %	44,81%
M	69,77 %	49,55 %	55,19%
Totaal	100%	100%	100%

d. Ze zijn gelijk.

Opgave 82

a. Kwalitatieve variabelen.

b. Bij *geslacht* niet; bij *plezier* wel.



Opgave 83

- Profiel; nominaal.
- Verticaal percenteren, want je wilt *per profiel* weten hoe de cijfers verdeeld zijn.
-

Aantallen leerlingen	Profiel				Eindtotaal
	CM	EM	NG	NT	
Wiskundecijfer					
5	9,1%	2,4%	1,7%		1,9%
6	54,5%	14,3%	8,3%	7,3%	13,0%
7		42,9%	38,3%	17,1%	31,2%
8	27,3%	31,0%	36,7%	36,6%	34,4%
9			10,0%	36,6%	13,6%
10		4,8%			1,3%
(leeg)	9,1%	4,8%	5,0%	2,4%	4,5%
Totaal	100%	100%	100%	100%	100%

- De verdelingen van de cijfers over EM- en NG-profielen lijken redelijk op elkaar; bij het CM-profiel scoren de meesten een wat lager cijfer, terwijl de meesten in het NT-profiel een wat hoger cijfer scoren.

Opgave 84

Horizontaal percenteren:

Aantallen leerlingen	Profiel				Eindtotaal
	CM	EM	NG	NT	
Wiskundecijfer					
5	33,3%	33,3%	33,3%		100%
6	30,0%	30,0%	25,0%	15,0%	100%
7		37,5%	47,9%	14,6%	100%
8	5,7%	24,5%	41,5%	28,3%	100%
9			28,6%	71,4%	100%
10		100,0%			100%
(leeg)	14,3%	28,6%	42,9%	14,3%	100%
Eindtotaal	7,1%	27,3%	39,0%	26,6%	

Vijven worden gelijkmatig over de niet-NT-profielen verdeeld; zessen relatief gelijkmatig over alle vier de profielen (maar iets minder in de N-profielen); zevens vooral in EM/NG; achten vooral in het NG-profiel, maar ook in alle andere profielen en negens alleen in de N-profielen.

Opgave 85

- ICT.
- Eigen antwoord.

Opgave 86

Verticaal percenteren: zelf doen.

Opgave 87

a.

	Kleurenblind	Niet kleurenblind	Totaal
Man	65	535	600
Vrouw	7	393	400
Totaal	72	928	1000

b. $\frac{65}{72}$ ofwel 90%.

c. $\frac{65}{600}$ ofwel 11%.

d. Vermoedelijk is er wel een statistisch verband: bij mannen is er een grotere kans op kleurenblindheid.

Opgave 88

Als het een warme, zonnige zomer is, dan worden er meer ijsjes verkocht en tevens meer zonnebrillen. Evenzo, bij een slechte zomer zullen er minder ijsjes en minder zonnebrillen worden verkocht. Er is geen direct verband, want beiden zijn een gevolg van een andere 'onzichtbare' variabele (*gemiddelde zomertemperatuur of aantal zonuren*).

Opgave 89

a. (Ongeveer) 20%.

b. 100%.

c. Die zit tussen 5% en 10% (?).

d. Die blijft dalen en gaat richting de 0% (maar zal nooit 0% worden).

Opgave 90

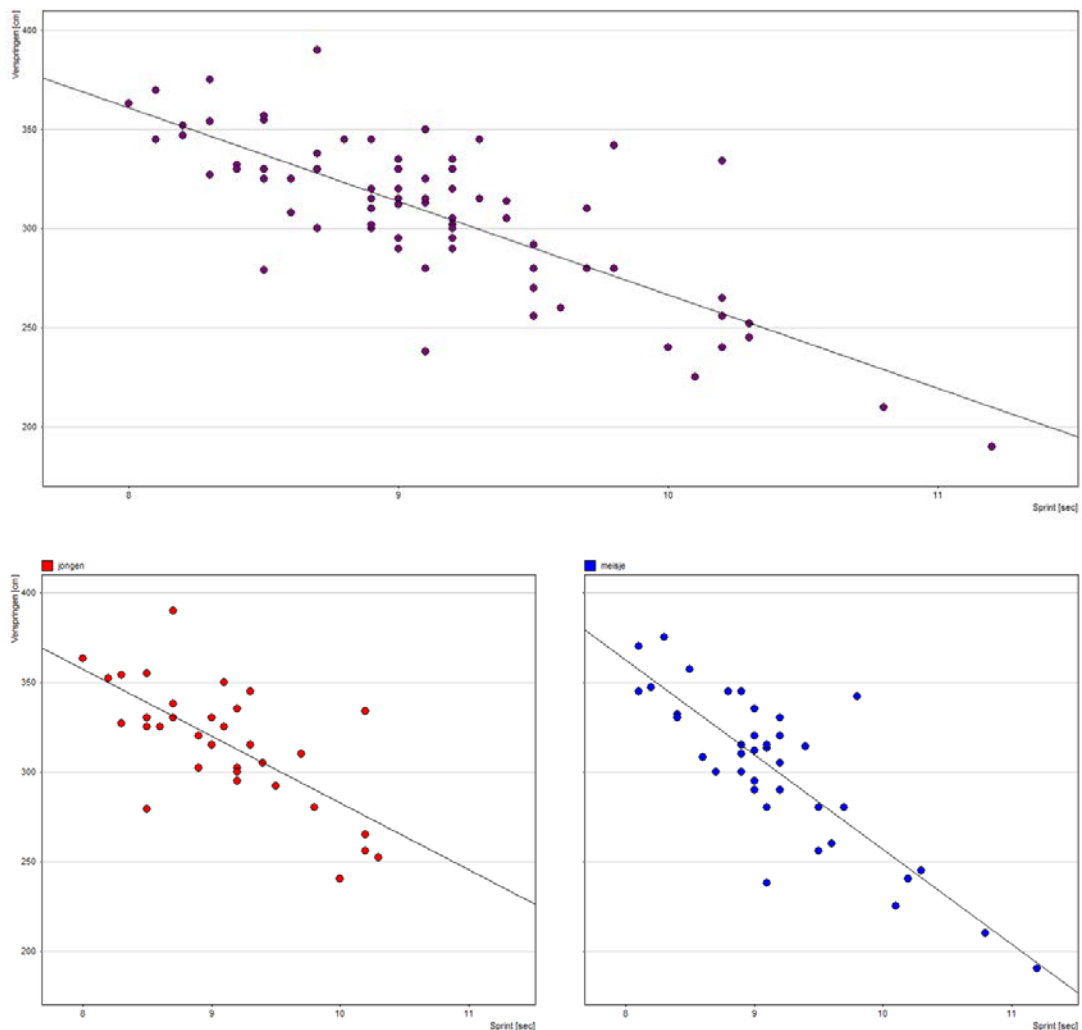
a. Horizontaal.

b.

Geslacht	Verspringen [cm]					Totaal
	150-199	200-249	250-299	300-349	350-399	
Jongen	0%	3,03%	21,21%	57,58%	18,18%	100%
Meisje	2,50%	12,50%	25,00%	52,50%	7,50%	100%
Totaal						



C.



Er is een verband en het is duidelijker als je de data splitst op geslacht.

Opgave 91

- In Noorwegen zijn er meer vuurwapens in omloop dan in Frankrijk en er zijn minder sterfgevallen door vuurwapens dan in Frankrijk; en zo zijn er veel meer landen die onder de getekende trendlijn liggen: daar zijn minder of evenveel sterfgevallen door vuurwapens als in landen met minder vuurwapens in omloop.
- Op Finland na zijn er in de VS de meeste vuurwapens in omloop en daar zijn enorm veel meer sterfgevallen door vuurwapens dan in alle landen die minder vuurwapens in omloop hebben.
-

Opgave 92

- Als de diameter van een boom toeneemt (de boom wordt dikker), neemt zijn volume toe.
- Kijk bij diameter 30 (inch) op de rechte lijn (door extrapoleren): volume is ongeveer 117 ft².
- Diameter is ongeveer 15,7 inch. Dan is volume ongeveer 43 ft², dat is ongeveer 40.000 cm².