

## 2.3 Verdelingen typeren

### Inleiding

Als je frequentieverdelingen bij datasets in beeld brengt, dus in diagrammen verwerkt, krijg je soms mooie symmetrische plaatjes. Maar lang niet altijd: sommige verdelingen zijn scheef, sommige erg grillig. Maar ook kun je met meerdere toppen te maken hebben.

#### Je leert in dit onderwerp

- de vorm van frequentieverdelingen te typeren;
- een cumulatief frequentiepolygoon te gebruiken en te interpreteren.

#### Voorkennis

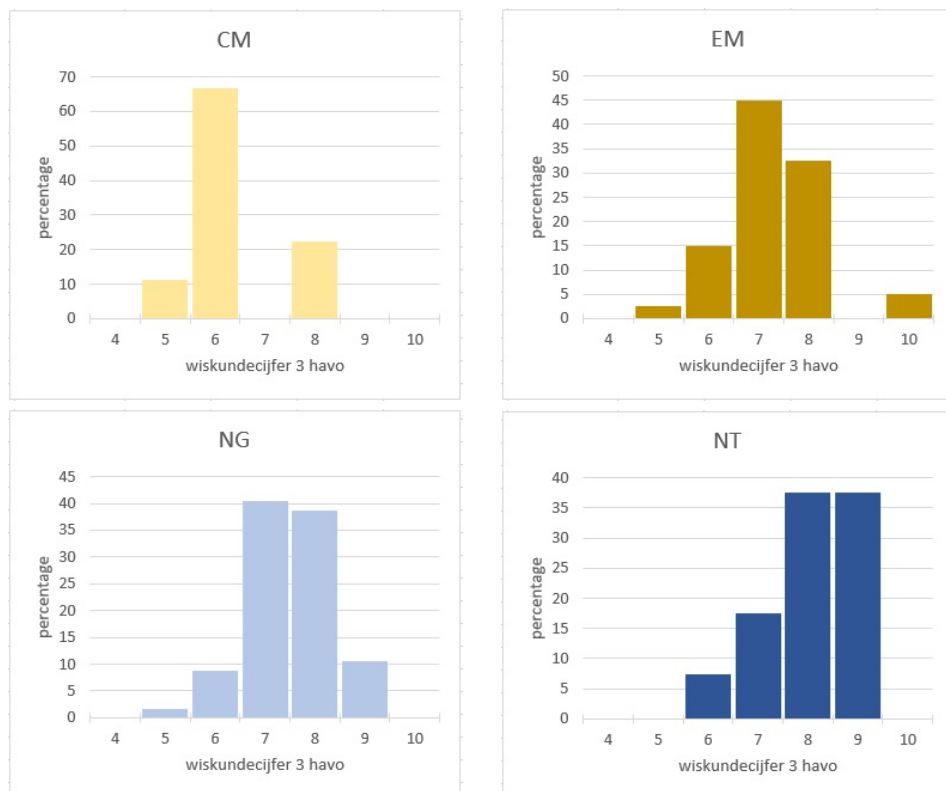
- de begrippen data, populatie, steekproef, aselekt en representatief, kwantitatief en kwalitatief, absolute en relatieve frequentie, discrete en continue variabele, klassenbreedte, klassenmidden en klassengrens;
- de verschillende centrummaten en spreidingsmaten berekenen;
- allerlei diagrammen interpreteren.

### Verkennen

#### Opgave V1

Gebruik weer de dataset **Gegevens 154 havo 4-leerlingen**.

Je ziet diagrammen met de wiskundecijfers in 3 havo voor de leerlingen in de verschillende profielen. Kijk per profiel naar de verdeling van de cijfers voor wiskunde in 3 havo.



Figuur 1

- a Wat valt je op aan de verschillen in vorm van deze frequentieverdelingen? Beschrijf je opmerkingen per profiel.

- b** Bij welke van deze vier verdelingen liggen de mediaan en het gemiddelde het dichtst bij elkaar? Hoe zie je dat aan het diagram?
- c** De verdeling van de wiskundecijfers in havo 3 voor het NT-profiel is een scheve verdeling. Waarom zou je deze verdeling scheef noemen, denk je?
- d** Welke verdeling is het meest symmetrisch?

### Opgave V2

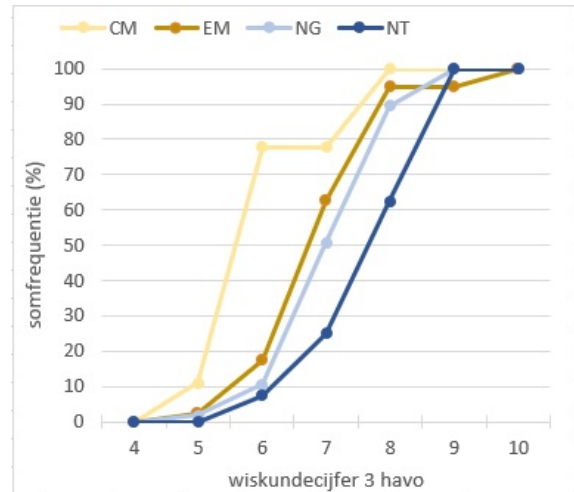
Bekijk de verdeling van het profiel NG nog eens. Ga ervan uit dat alle frequenties hele percentages zijn.

- a** Hoeveel procent van de NG-leerlingen had in 3 havo een wiskundecijfer lager dan of gelijk aan 7?
- b** Hoeveel procent van die leerlingen had in 3 havo een wiskundecijfer lager dan of gelijk aan 8?

Je noemt de frequenties die je bij a en b hebt berekend wel 'somfrequenties'. Bij somfrequenties tel je bij de frequentie van bijvoorbeeld het cijfer 7 ook de frequenties van alle voorgaande cijfers op: je stapelt de frequenties als het ware op elkaar. In dit geval stapel je relatieve frequenties op elkaar.

- c** Maak een tabel waarin bij elk cijfer de relatieve somfrequentie staat.
- d** Welk percentage had een cijfer lager dan of gelijk aan 5?
- e** Bekijk het lijndiagram.

Welke conclusie kun je uit dit diagram trekken?



Figuur 2

### Uitleg

Bij een statistische variabele zijn veel verschillende verdelingen mogelijk.

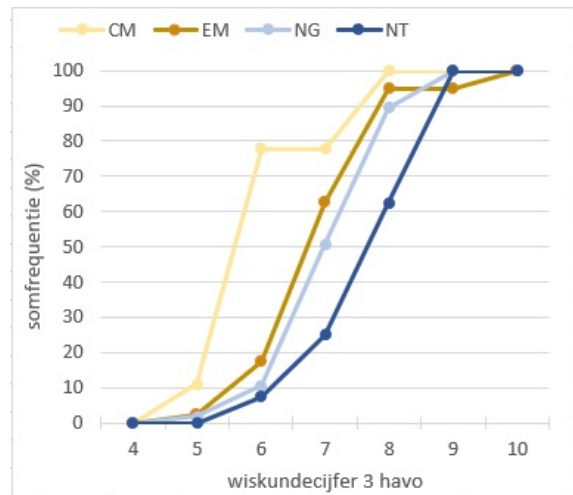
Als je frequentieverdelingen bij datasets in beeld brengt, dus in diagrammen verwerkt, krijg je soms mooie symmetrische plaatjes, maar lang niet altijd. Sommige verdelingen zijn scheef, sommige erg grillig. Maar ook kun je met meerdere toppen te maken hebben.

Let bijvoorbeeld eens op de vorm en de verdeling van een staafdiagram. Is het diagram scheef of juist symmetrisch? Zijn er meerdere toppen of is er juist sprake van een opvallende gelijkmatigheid? Zijn er veel uitschieters?

Een frequentieverdeling kun je ook in beeld brengen met behulp van somfrequenties: dat is de totale frequentie voor een waarde opgeteld bij alle waarden daaronder. Je stapelt de frequenties dus op elkaar: bij elke frequentie tel je die van de voorgaande waarden (of klassen) op. Je noemt dit de cumulatieve frequenties ('cumuleren' betekent 'opstapelen').

Je kunt een lijndiagram maken van die somfrequenties (cumulatieve frequenties). Zo'n lijndiagram noem je een 'cumulatief frequentiepolygoon'.

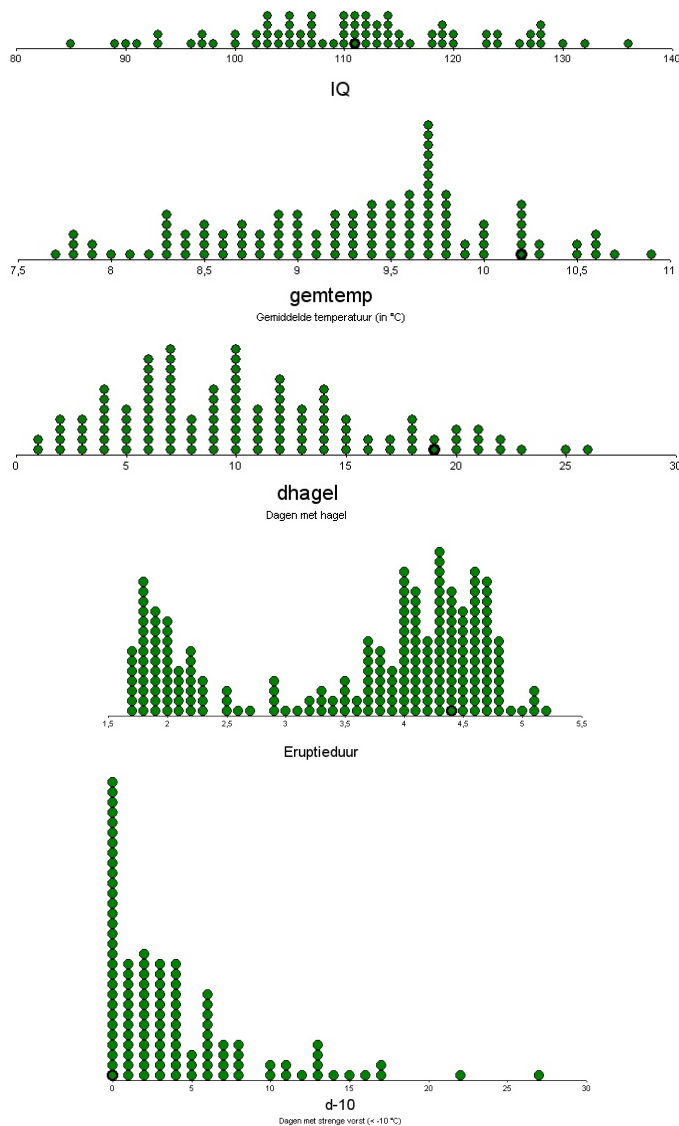
Als je te maken hebt met een klassenindeling, verbind je om een lijndiagram van de cumulatieve frequenties te krijgen, de frequenties van de rechterklassengrenzen met elkaar.



**Figuur 3** cumulatieve frequentiepolygoon

### Opgave 1

Bekijk de dotplots van een aantal datasets. Beschrijf van elke dataset de vorm van de verdeling. Is er sprake van symmetrie, een gelijkmatige verdeling, meerdere toppen of uitschieters?



**Figuur 4**

## Opgave 2

Je ziet diagrammen van de lengteverdeling van sporters. De lengtes zijn ingedeeld in klassen met een klassenbreedte van 5 cm. Eén ervan gaat over basketballers, één over hardlopers en één over gewichtheffers.

- Welk diagram is het meest gelijkmatig, welk diagram heeft meerdere toppen en welk diagram vind je het meest symmetrisch?
- Welk diagram hoort bij de basketballers, welk bij de hardlopers en welk bij de gewichtheffers? Licht je antwoord toe.
- Bij welk van deze diagrammen zitten de mediaan en het gemiddelde beide ongeveer in het midden van de verdeling?

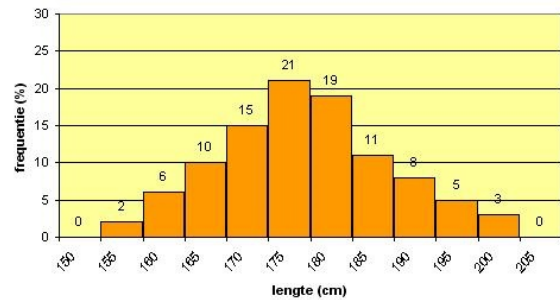


diagram I

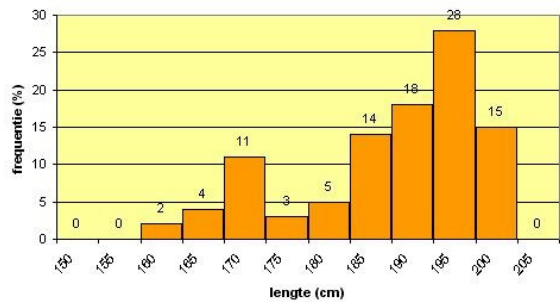


diagram II

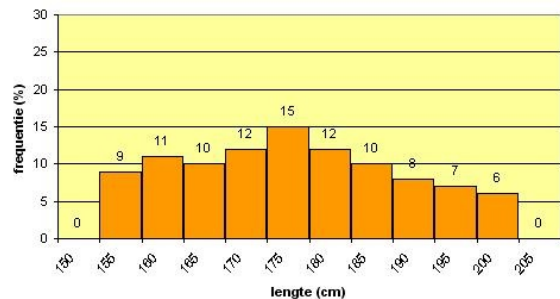


diagram III

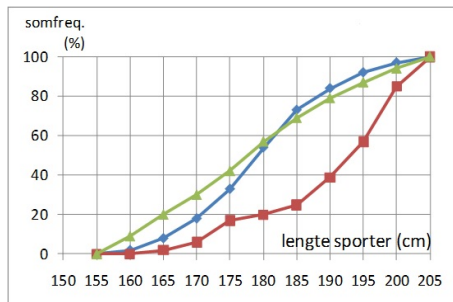
Figuur 5

## Opgave 3

Je kunt de lengteverdelingen in de vorige opgave ook vergelijken met behulp van somfrequentiepolygonen of cumulatieve frequentiepolygonen.

- Welke klassenindeling is er bij de diagrammen gebruikt? Wat zijn de klassenmiddens?
- Neem diagram I. Hoeveel procent van de sporters is kleiner dan 170 cm?
- Maak een tabel van de klassen bij diagram I met daarbij de cumulatieve frequenties.
- Waarom moet je bij klassen de somfrequenties boven de rechterklassengrenzen uitzetten?

- e Je ziet drie somfrequentiepolygonen (cumulatieve frequentiepolygonen) die bij de drie diagrammen horen. Geef aan welk diagram bij welke kleur hoort.

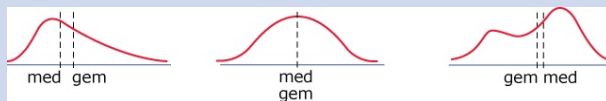


Figuur 6

- f Wat kun je uit de cumulatieve frequentiepolygonen aflezen over de lengteverdelingen van de sporters?

## Theorie en voorbeelden

### Om te onthouden



Figuur 7

Bij de **vorm van een frequentieverdeling** let je op:

- de symmetrie
- de scheefheid
- het aantal toppen
- de uitschieters
- de gelijkmatigheid

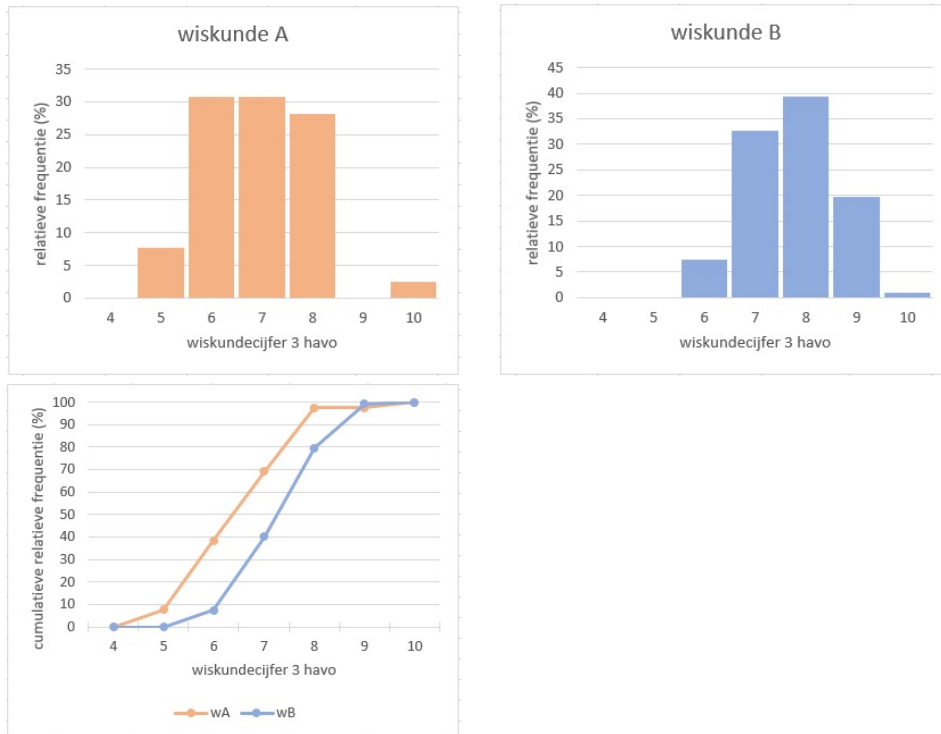
Bij een symmetrische verdeling vallen mediaan (med) en gemiddelde (gem) vrijwel samen. Dat zie je in het middelste diagram. Je ziet ook dat daarvan het aantal uitschieters in de staart links en rechts ongeveer gelijk is.

De **somfrequentie** of **cumulatieve frequentie** is de totale frequentie van een waarde plus die van alle kleinere waarden. Je stapelt de frequenties op elkaar: bij elke frequentie tel je de voorgaande frequenties op. Een lijndiagram van de somfrequenties van een verdeling noem je een **somfrequentiepolygoon** of **cumulatief frequentiepolygoon**. Bij een klassenindeling worden de somfrequenties bepaald door de frequenties van een klasse bij alle voorgaande klassen op te tellen. Deze somfrequenties zet je in een diagram uit boven de rechterklassengrenzen.

## Voorbeeld 1

Gebruik weer de dataset **Gegevens 154 havo 4-leerlingen**.

Je wilt nagaan of leerlingen die wiskunde B kiezen, beter waren in wiskunde in de onderbouw dan leerlingen die wiskunde A kiezen. Daartoe bekijk je de variabele cijfers (het eindcijfer voor wiskunde in havo 3) voor elk van deze deelgroepen.



**Figuur 8**

Het linker diagram lijkt redelijk symmetrisch met als top het cijfer 8. De mediaan van deze gegevens is 8 en het gemiddelde is 7,7. Het rechter diagram is in het midden meer gelijkmatig en er is geen echte top. De mediaan van deze gegevens is 7 en het gemiddelde 6,9. Het cijfer 10 wijkt behoorlijk veel af van de andere cijfers, maar is nog net geen uitschieter.

Bekijk de somfrequentiepolygonen van beide verdelingen.

Je ziet dat de wiskunde B-leerlingen stelselmatig hogere cijfers hebben (op de uitschieter na). Ongeveer 40% van de A-leerlingen had bijvoorbeeld een wiskundecijfer 6 of lager, tegen nog geen 10% van de B-leerlingen.

## Opgave 4

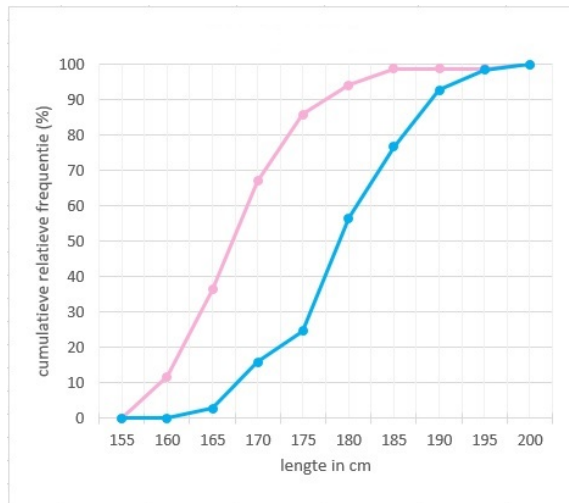
Bekijk **Voorbeeld 1**. Alle percentages zijn gehele getallen.

- Waarom zijn geen van beide verdelingen scheef?
- Reken de gemiddelden en de medianen van beide verdelingen na.
- De somfrequenties zijn uitgezet tegen de gehele cijfers 5, 6, 7, 8, 9, 10. Is dat hier correct?
- Hoeveel procent van de leerlingen met wiskunde A heeft een cijfer 7 of lager? En hoeveel procent van de leerlingen met wiskunde B heeft zo'n cijfer? Kun je nu iets zeggen over het aantal leerlingen met een 7 of lager?

## Opgave 5

Bekijk de frequenties (in procenten) van de lengtes van de 154 meisjes en jongens in een brugklas.

klassen			meisjes		jongens	
onder	midden	boven	freq	%	freq	%
150	152,5	155	0	0	0	0
155	157,5	160	10	12	0	0
160	162,5	165	21	25	2	3
165	167,5	170	26	31	9	13
170	172,5	175	16	19	6	9
175	177,5	180	7	8	22	32
180	182,5	185	4	5	14	20
185	187,5	190	0	0	11	16
190	192,5	195	0	0	4	6
195	197,5	200	1	1	1	1
			85	100	69	100



Figuur 9

- Welke somfrequentiepolygoon hoort bij de meisjes en welke bij de jongens? Licht je antwoord toe.
- Welke conclusie kun je uit dit cumulatief frequentiepolygoon trekken?

## Voorbeeld 2

Bekijk het cumulatieve relatieve frequentiepolygoon van deze klassenindeling van levendgeborenen naar leeftijd van de moeder. Je ziet hoe hierbij een passend boxplot is gemaakt.

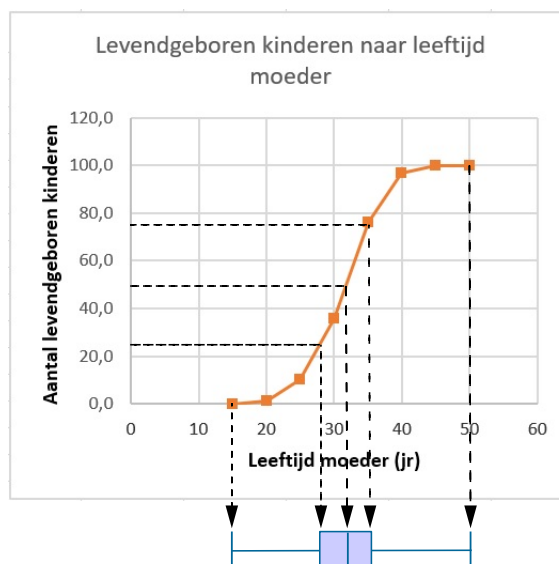
Zoek voor de boxplot de vijf getallen om de frequentieverdeling samen te vatten: minimum,  $Q_1$ ,  $Q_2$  (mediaan),  $Q_3$  en maximum.

Antwoord

Lees bij 50% de mediaan af, bij 25% het eerste kwartiel en bij 75% het derde kwartiel. Minimum en maximum zitten bij 0% en bij 100%.

Je ziet:

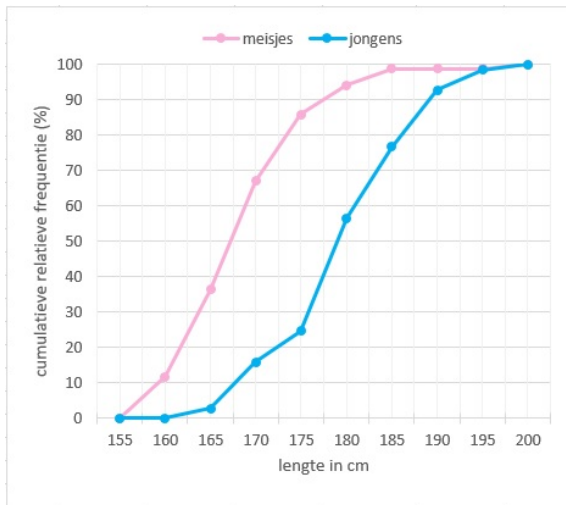
- het minimum is 15
- het eerste kwartiel  $Q_1$  is 28
- de mediaan  $Q_2$  is 31,5
- het derde kwartiel is 35
- het maximum is 50



Figuur 10

## Opgave 6

Bepaal vanuit deze cumulatieve frequentiepolygonen de vijf getallen om beide frequentieverdelingen afzonderlijk samen te vatten in een boxplot.

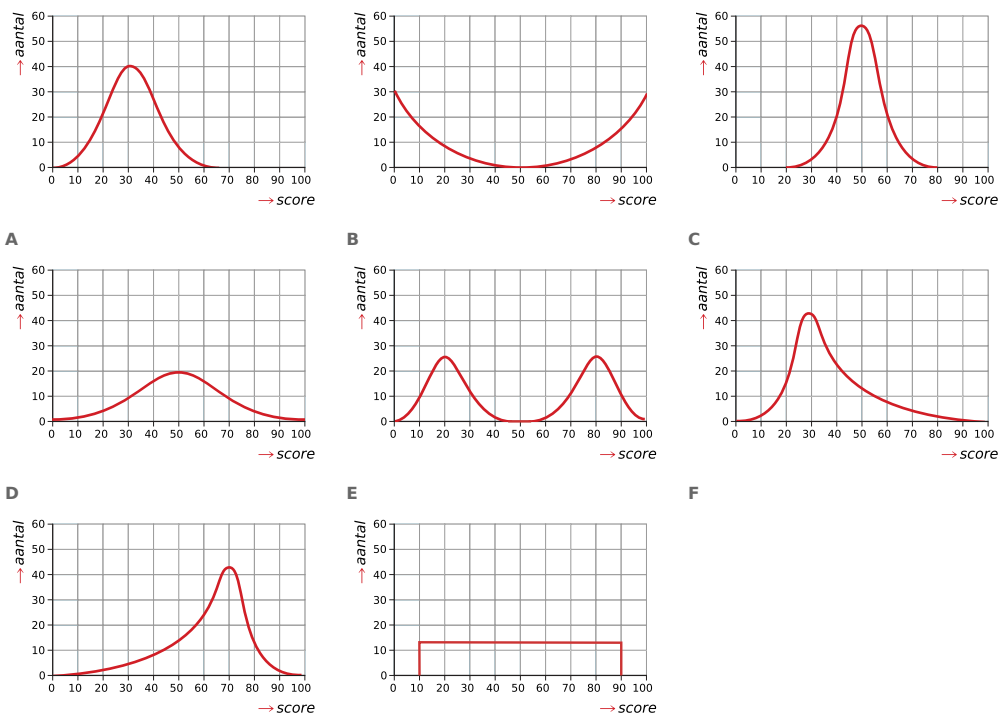


Figuur 11

## Verwerken

## Opgave 7

Bekijk deze frequentieverdelingen.



G

H

Figuur 12

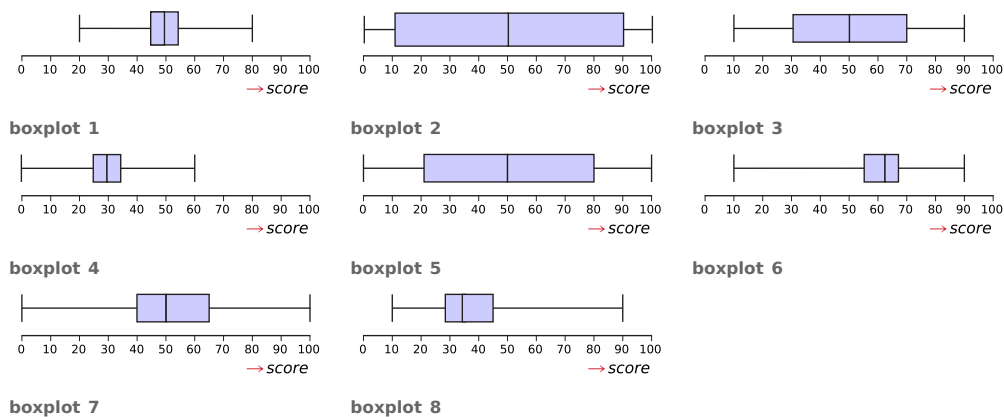
- Welke van deze verdelingen zijn symmetrisch?
- Welke van deze verdelingen zijn eentoppig?
- Welke van deze verdelingen zijn links scheef?

(naar: Mathematics Assessment Project, Classroom challenges, maart 2015)



## Opgave 8

Bekijk deze frequentieverdelingen in de vorm van boxplots.



Figuur 13

- Welke van deze verdelingen zijn symmetrisch?
- Welke van deze verdelingen zijn links scheef?

(naar: Mathematics Assessment Project, Classroom challenges, maart 2015)

## Opgave 9

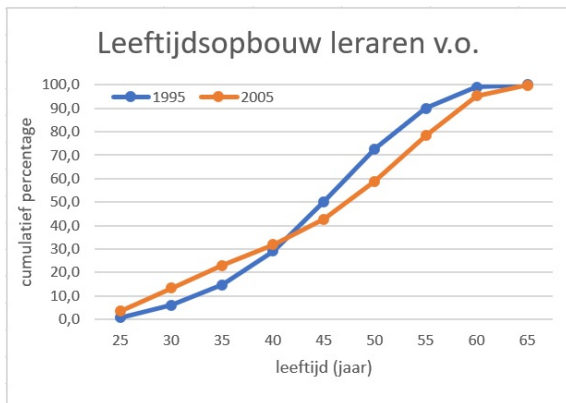
Bekijk de relatieve frequenties bij de klassenverdeling van de leeftijdsopbouw van leraren in het primair onderwijs en het voortgezet onderwijs in de jaren 1995 en 2005.

Leeftijdsopbouw leraren				
leeftijd	po		vo	
	1995	2005	1995	2005
20 - 24	3,7	6,1	0,7	3,3
25 - 29	11,9	15,6	5,3	9,9
30 - 34	12,3	11,1	8,7	9,7
35 - 39	18,2	8,1	14,5	8,8
40 - 44	21,3	10,6	21,0	10,9
45 - 49	18,3	16,6	22,5	16,3
50 - 54	9,5	16,7	17,3	19,4
55 - 59	4,4	12,7	9,1	17,1
60 - 64	0,5	2,5	1,0	4,5
	100,0	100,0	100,0	100,0

Tabel 1

- Maak staafdiagrammen voor de leeftijdsopbouw in 1995 en in 2005 van de leraren in het p.o. en die in het v.o. Beschrijf de verschillen tussen beide verdelingen.
- Leg uit waar je de nieuwe instroom van leraren in het primair onderwijs aan herkent.
- Bepaal de klassenmiddens en geef daarmee een schatting van de gemiddelde leeftijden in het primair onderwijs in 1995 en in 2005.

- d Bekijk de cumulatieve relatieve frequentiepolygonen voor het voortgezet onderwijs. Lees de kwartielen af voor 2005.

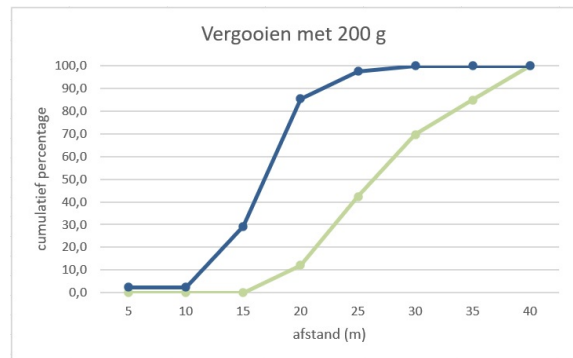


Figuur 14

### Opgave 10

Je ziet de cumulatieve frequentieverdelingen van de prestaties van 74 brugklassers met vergooien.

- a Vergelijk de twee cumulatieve relatieve frequentieverdelingen. De jongens bleken verder te kunnen gooien dan de meisjes. Welke is van de jongens en welke van de meisjes denk je? Licht je antwoord toe. En wat is de minst verre worp van de meisjes en wat is de minst verre worp van de jongens?
- b Je wilt een uitspraak doen over de spreiding van de verdelingen aan de hand van de cumulatieve relatieve frequentiepolygonen. Bepaal voor de meisjes en de jongens met behulp van de grafiek de mediaan, de spreidingsbreedte en de kwartielaafstand. Wat kun je concluderen uit deze spreidingsmaten?

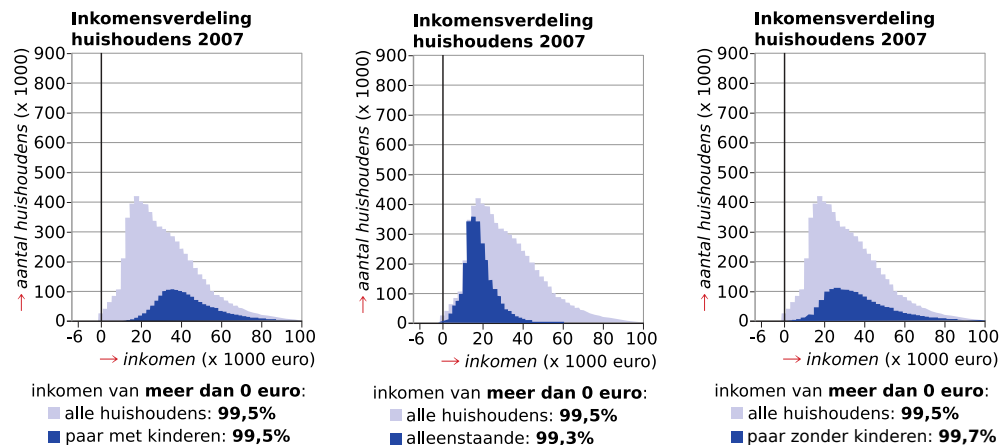


Figuur 15

### Opgave 11

Je ziet drie staafdiagrammen die de inkomensverdeling in Nederland in 2007 in kaart brengen.

Het gestandaardiseerd inkomen is het besteedbaar inkomen gecorrigeerd voor verschillen in grootte en samenstelling van het huishouden. Dus alle inkomens worden herleid tot het inkomen van een eenpersoonshuishouden.



Figuur 16

- a Beschrijf de vormen van al deze verdelingen.

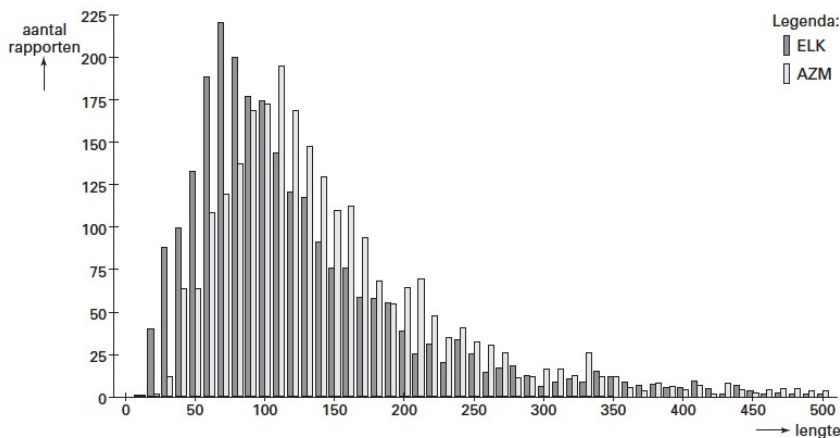
- b** Bepaal voor elke verdeling de spreidingsbreedte.
- c** Ga voor elke verdeling na hoe het modale inkomen, de mediaan en het gemiddelde ten opzichte van elkaar liggen.
- d** Vergelijk de inkomensverdelingen van de deelgroepen 'paar met kinderen', 'paar zonder kinderen' en 'alleenstaanden'. Probeer ook een verklaring voor de verschillen te geven.

(bron: CBS)

## Toepassen

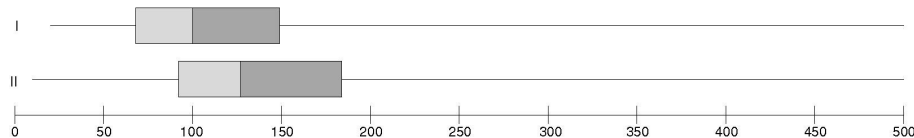
### Opgave 12: Medische rapporten

In ziekenhuizen worden vaak medische rapporten geschreven. Bij een onderzoek naar de inhoud van dergelijke rapporten zijn 2500 rapporten van het Elkerliek Ziekenhuis (ELK) in Deurne vergeleken met 2500 rapporten van het Academisch Ziekenhuis Maastricht (AZM). Van elk rapport is de lengte bepaald; de lengte van een rapport is het aantal woorden dat het bevat. In de figuur zijn de gegevens weergegeven in een gecombineerd staafdiagram met klassenbreedte 10.



Figuur 17

- a** Beschrijf de overeenkomst in de vorm van de twee verdelingen.
- b** Welke van deze boxplots, I of II, hoort bij de rapporten van het ELK? Licht je antwoord toe.



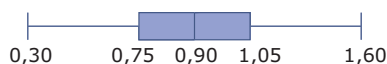
Figuur 18

- c** Kun je op grond van de twee boxplots concluderen dat er een verschil is tussen de lengtes van de rapporten in de twee ziekenhuizen? Beargumenteer je antwoord.
- d** Uit het onderzoek bleek dat de mediaan en het gemiddelde die horen bij de rapporten van het AZM niet even groot zijn. Geef met een redenering, dus zonder een berekening, aan of de mediaan groter of kleiner is dan het gemiddelde.

(bron: pilotexamen wiskunde A in 2004, tweede tijdvak)

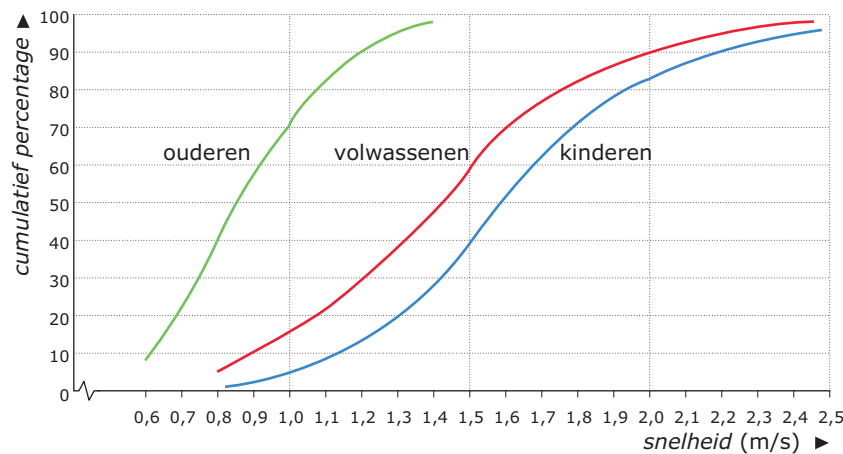
### Opgave 13: Loopsnelheden van voetgangers

Men heeft onderzoek gedaan naar de loopsnelheden van voetgangers. Bij dit onderzoek zijn de voetgangers in drie leeftijdsgroepen verdeeld, namelijk kinderen, volwassenen en ouderen. Met de gegevens uit het onderzoek heeft men een boxplot gemaakt voor de loopsnelheden van de groep ouderen.



Figuur 19

De snelheden die bij de boxplot vermeld zijn, zijn in meters per seconde. Meer gedetailleerde informatie over de groepen zie je in de figuur.

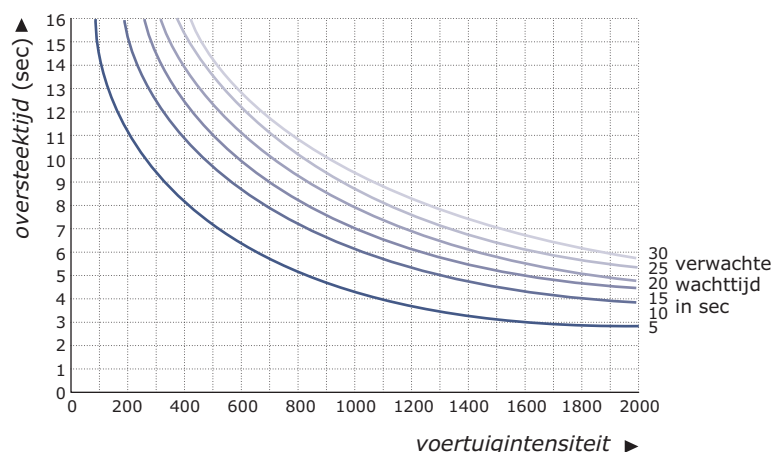


**Figuur 20**

Op de verticale as staat een cumulatief percentage; dit houdt in dat afgelezen kan worden hoeveel procent van de mensen van de verschillende groepen loopt met de aangegeven snelheid of een lagere snelheid. Zo kun je aflezen dat voor de groep ouderen bij een snelheid van 1 m/s het cumulatieve percentage 80 is. Dus bijna 80% van de ouderen loopt met een snelheid van 1 m/s of langzamer. Aan de hand van onder andere deze gegevens wordt een model gemaakt voor de tijd die mensen nodig hebben om een weg over te steken. Neem aan dat de loopsnelheden ook voor het oversteken van een weg gelden. We bekijken het oversteken van een twintig meter brede weg. Er wordt recht overgestoken, dus men loopt daarbij twintig meter.

- Leg uit hoe de boxplot voor de oversteektijden van ouderen uit de grafiek is ontstaan.
- Kun je de vorm van de verdeling van de oversteektijd bij ouderen typeren?

Tot nu toe heb je alleen gekeken naar de tijd van het oversteken zelf. Als je bij een weg aankomt, kun je niet altijd meteen oversteken; soms moet je een aantal seconden wachten. Deze wachttijd hangt samen met de drukte op de weg en de benodigde oversteektijd. De drukte op de weg wordt aangegeven met het aantal voertuigen dat per uur passeert (voertuigintensiteit). Omdat ouderen in het algemeen minder snel lopen, zullen voor deze groep de benodigde oversteektijd en dus ook de wachttijd groter zijn dan bijvoorbeeld voor kinderen. Er is een model gemaakt voor de samenhang tussen oversteektijd, voertuigintensiteit en verwachte wachttijd.



**Figuur 21**

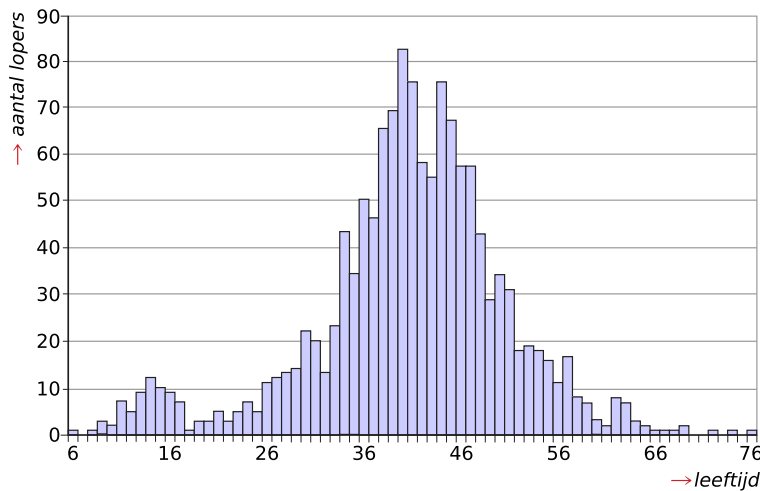
- Leid de grafiek af die het verband aangeeft tussen de oversteektijd en de verwachte wachttijd bij een voertuigintensiteit van 800. Noem de coördinaten van minimaal vijf punten van de grafiek voor wachttijden van 5 tot en met 30 seconden.

(naar: examen wiskunde A in 1994, tweede tijdvak)

## Testen

### Opgave 14

Je ziet de leeftijden van de deelnemers aan een hardloopwedstrijd in 2010.

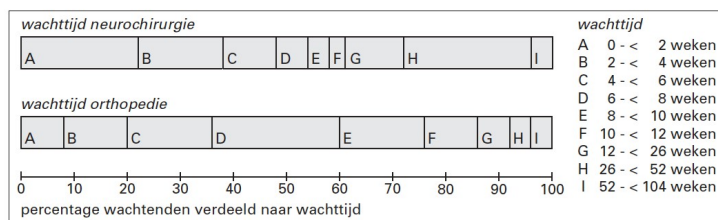


Figuur 22

- Beschrijf de vorm van deze verdeling. Gebruik woorden als symmetrie, scheefheid, toppen, staart, uitschieters en gelijkmatigheid.
- Hoe kun je zien dat het niet om een cumulatieve frequentie (sommfrequentie) gaat?
- Wat kun je zeggen over de centrummaten bij deze verdeling? En over de spreidingsmaten?
- Welke leeftijdsgroepen zullen wat meer gestimuleerd moeten worden om mee te doen, kijkend naar deze verdeling?

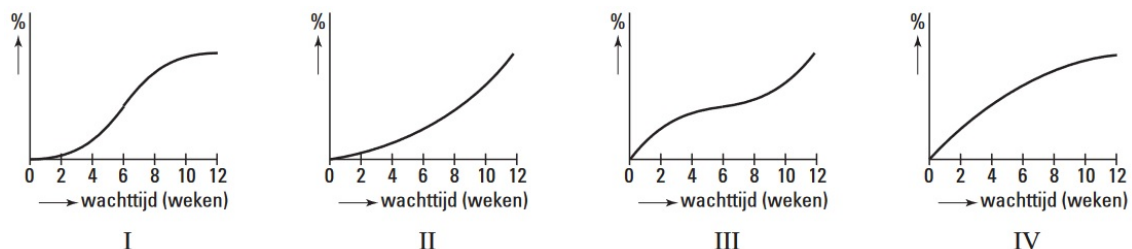
### Opgave 15

De wachtlijsten voor neurochirurgie en orthopedie zijn in kaart gebracht. Je ziet hoeveel mensen op een wachtlijst staan en hoelang ze moeten wachten.



Figuur 23

Je ziet vier cumulatieve frequentiepolygonen (I, II, III en IV) voor de eerste twaalf weken.



Figuur 24


Welke van deze vier cumulatieve frequentiepolygonen past het best bij de wachttijden tot twaalf weken bij neurochirurgie? Licht je antwoord toe.

(naar: examen havo wiskunde A in 2003, tweede tijdvak)



© 2024

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt. Klik op  in de marge bij de betreffende opgave. Uw mailprogramma wordt dan geopend waarbij het emailadres en onderwerp al zijn ingevuld. U hoeft alleen uw opmerkingen nog maar in te voeren.

Email: [f.spijkers@math4all.nl](mailto:f.spijkers@math4all.nl)

Met de Math4All Foliostraat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij [a.f.otten@math4all.nl](mailto:a.f.otten@math4all.nl) een gratis inlog voor de maatwerkdienst aanvragen.

---