

## 3.4 Data samenvatten

### Inleiding

Grote verzamelingen statistische gegevens zijn (ondanks het gebruik van tabellen en diagrammen) vaak nogal onoverzichtelijk. Vergelijken van frequentieverdelingen is ook niet altijd in één oogopslag mogelijk. Daarom worden de resultaten van een statistisch onderzoek vaak samengevat in een aantal getallen die snel informatie geven over frequentieverdelingen, zoals gemiddelden en spreiding van gegevens.

Een aantal opgaven uit dit onderdeel is afkomstig uit de NLT-module: 'Maak het verschil', net als enkele datasets.

#### Je leert in dit onderwerp

- een reeks waarnemingen samen te vatten met centrummaten: gemiddelde, mediaan en modus;
- een reeks waarnemingen samen te vatten met spreidingsmaten: variantie, spreidingsbreedte, kwartielf afstand en standaardafwijking;
- een reeks waarnemingen samen te vatten in een boxplot.

#### Voorkennis

- gegevens in een frequentietabel (met zowel absolute als relatieve als cumulatieve frequenties) verwerken;
- een (cumulatief) histogram en een (cumulatieve) frequentiepolygoon maken bij een frequentietabel.

### Verkennen

#### Opgave V1

De schoolleiding en de ouderraad willen een goed beeld krijgen van de aanwezigheid van leerlingen in jullie leerjaar bij de diverse vakken. Ze willen dat in een beknopt en helder verslag van maximaal twee A4'tjes uiteengezet hebben.

Hoe kan zo'n beknopt verslag eruit zien, zodat iedereen goed geïnformeerd wordt? Denk aan het samenvatten van gegevens met behulp van diagrammen, gemiddelden en uitersten.

#### Uitleg

Jouw klas heeft een toets gehad. Je docent doet de mededeling dat de toets goed is gemaakt met een gemiddelde van 7,3. Ben je blij met deze informatie of hoor je liever dat het modale cijfer 7,3 is? Of dat de mediaan 7,3 is?

Met deze mededeling probeert je docent een frequentieverdeling met één getal te karakteriseren.

- Het modale cijfer is het cijfer dat het vaakst voorkomt. Hier zegt het niet veel, want misschien komt 7,3 twee keer voor en zijn alle andere cijfers heel verschillend.
- Ook een mediaan (middelste cijfer) van 7,3 zegt niet veel, hoewel je dan zeker weet dat de helft van de cijfers hoger dan of gelijk aan 7,3 is (en de andere helft lager dan of gelijk aan 7,3).
- Het gemiddelde krijg je door alle cijfers op te tellen en te delen door het totale aantal leerlingen. Maar ben jij een gemiddelde of bovengemiddelde leerling?

Deze getallen zeggen op zichzelf weinig. Het wordt al beter als je er een mededeling over de spreiding van de cijfers bij krijgt. Het gemiddelde cijfer is een 7,3 en de cijfers hebben een spreiding van 2,2 bijvoorbeeld.

Maar wat wordt onder de spreiding verstaan? Het verschil tussen het hoogste en het laagste cijfer, de spreidingsbreedte, is bijvoorbeeld zo'n spreidingsmaat. Maar er zijn ook andere spreidingsmaten. Een boxplot, waarin onder andere mediaan en spreiding zijn verwerkt, kan je meer duidelijkheid verschaffen.

### Opgave 1

Bekijk de **Uitleg**. In de tabel zie je de cijfers van een wiskundetoets van twee parallelklassen.

cijfers klas A					cijfers klas B				
6,7	6,4	4,9	3,8	4,0	4,0	6,2	4,9	3,9	5,9
5,6	5,8	6,8	8,2	4,7	7,3	4,7	6,7	7,6	9,4
3,4	8,5	4,1	6,9	7,3	8,3	5,7	7,2	8,7	7,1
6,1	7,5	6,7	6,2	3,4	7,0	6,5	7,4	5,0	4,8
7,9	4,5	8,3			7,7	6,5	4,9	8,8	6,3

Tabel 1

- a Waarom heeft het geen zin om van beide klassen het modale cijfer te vergelijken?
- b Bepaal van beide klassen de mediaan.
- c Zegt de mediaan iets over welke klas beter heeft gescoord?
- d Bereken van beide klassen het gemiddelde cijfer.
- e Welke van beide klassen heeft het hoogste gemiddelde? Kun je nu zonder meer zeggen dat die klas ook beter heeft gescoord?

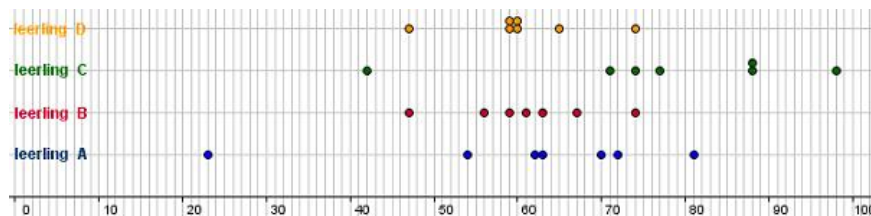
### Opgave 2

Je ziet de SE-cijfers (schoolexamen) van enkele leerlingen aan het eind van havo 5. Hun eindcijfer SE is het gemiddelde van deze cijfers.

	se1	se2	se3	se4	se5	se6	se7
A	7,2	6,3	7,0	2,3	6,2	8,1	5,4
B	5,9	7,4	5,6	6,7	6,1	6,3	4,7
C	8,8	9,8	7,4	8,8	5,7	7,3	4,2
D	7,5	6,0	6,0	6,5	6,1	6,1	4,7

Figuur 1

Elk SE-cijfer telt even zwaar mee. In de figuur is voor elke leerling elk SE-cijfer aangegeven door een bolletje op een getallenlijn (de komma in het cijfer is weggelaten).



Figuur 2

- a De leerlingen A en B hebben hetzelfde gemiddelde. Toch is hun cijferbeeld nogal verschillend. Hoe komt dat?
- b De spreiding van de cijfers van leerling A en C is vrijwel hetzelfde. Waarin verschilt hun cijferbeeld vooral?
- c De cijfers van de leerlingen B en D hebben dezelfde spreidingsbreedte. Is de spreiding van hun cijfers ook hetzelfde?

Een andere maat voor de spreiding vind je door te kijken hoe ver elk cijfer van het gemiddelde af ligt. Bereken van elk cijfer het verschil met het gemiddelde. Je ziet die verschillen voor leerling A.

	se1	se2	se3	se4	se5	se6	se7
A	7,2	6,3	7,0	2,3	6,2	8,1	5,4
	1,1	0,2	0,9	-3,8	0,1	2,0	-0,7

Figuur 3

- d Bereken het gemiddelde van deze verschillen. Verbaast het antwoord je? Licht je antwoord toe.

Het gemiddelde van deze verschillen is geen goede spreidingsmaat. Dat zit hem in de mintekens. Door te kwadrateren vallen die mintekens weg. Maak voor leerling A een lijst van de kwadraten van de verschillen.

- e Bereken daarvan het gemiddelde. Heb je nu een goede spreidingsmaat?  
Door het kwadrateren wordt het getal dat je bij e hebt gevonden, nogal groot. Dat los je op door de wortel uit dit getal te nemen. Je krijgt dan de standaardafwijking van de set cijfers.
- f Ga na dat de standaardafwijking voor leerling A ongeveer 1,73 is.
- g Bereken ook voor leerling B de verschillen van de cijfers met het gemiddelde. Bereken vervolgens het gemiddelde van de kwadraten van die verschillen en de standaardafwijking.

## Theorie en voorbeelden

### Om te onthouden

Getallen die (bij benadering) het midden aangeven van een reeks waarnemingen heten centrummaten. Er zijn drie **centrummaten**.

- De **modus** is de waarneming met de hoogste frequentie. Vooral geschikt voor kwalitatieve variabelen.
- De **mediaan** is het middelste waarnemingsgetal als de waarnemingsgetallen op volgorde van klein naar groot staan. Is het aantal even, dan zijn er twee middelste waarnemingsgetallen. De mediaan is dan het gemiddelde van die middelste twee.
- Het **gemiddelde** bereken je door alle waarnemingsgetallen op te tellen en te delen door het totale

aantal. Als je de waarnemingsgetallen  $x_1, x_2, x_3, \dots, x_n$  noemt, schrijf je dit als:  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ .

Daarin geldt  $\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$ .

De Griekse hoofdletter sigma ( $\Sigma$ ) is het somteken. Bij een frequentietabel vermenigvuldig je elk waarnemingsgetal met de frequentie.

Het gemiddelde is dan:  $\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{n}$ .

Bij **klassenindelingen** spreek je van de **modale klasse** en kun je de mediaan het beste opzoeken in een cumulatieve relatieve frequentiepolygoon (de waarde bij 50% schatten door aflezen). Het gemiddelde kun je dan alleen maar schatten door het gemiddelde van de **klassenmiddens** te berekenen.

Centrummaten alleen zeggen nog weinig, er hoort steeds een spreidingsmaat bij. Er zijn drie **spreidingsmaten**:

- De **spreidingsbreedte** (ook **variatiebreedte**) is het verschil tussen het hoogste en laagste waarnemingsgetal.
- De **interkwartielafstand** is het verschil tussen de mediaan van de grootste helft (het **derde kwartiel** of  $Q_3$ ) en de mediaan van de kleinste helft (het **eerste kwartiel** of  $Q_1$ ). Om de kwartielen te bepalen, zet je eerst de waarnemingsgetallen in volgorde van klein naar groot en verdeel je ze in twee helften. Bestaan de waarnemingen uit een oneven aantal waarden, dan wordt de mediaan van de hele set niet meegenomen om  $Q_1$  en  $Q_3$  te berekenen.
- De **standaardafwijking** (of **standaarddeviatie**) vind je door van elk waarnemingsgetal het verschil met het gemiddelde te bepalen en dat getal te kwadrateren. Die kwadraten tel je op en je deelt ze door het totale aantal waarnemingen. Dit getal heet de **variantie**. De wortel uit de va-

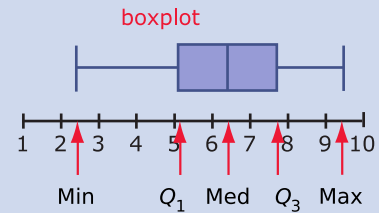
riantie is de standaarddeviatie  $\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{n}}$ . De Griekse (kleine) letter sigma is het teken voor standaardafwijking.

Bij klassenindelingen is de spreidingsbreedte het aantal klassen maal de klassenbreedte. De mediaan en de kwartielen zoek je het beste op in een cumulatieve relatieve frequentiepolygoon (de mediaan bij 50%, het eerste kwartiel bij 25% en het derde kwartiel bij 75%). De standaarddeviatie kun je nu

alleen schatten door de standaarddeviatie van de klassenmiddens te berekenen. De mediaan, het gemiddelde en alle spreidingsmaten kunnen alleen gebruikt worden voor kwantitatieve variabelen. Hoe je ze met de grafische rekenmachine bepaalt, zie je in het **Practicum**. Het is bij grotere datasets verstandiger om met Excel te werken.

De mediaan, het eerste en derde kwartiel en de spreidingsbreedte en de kwartielf afstand kun je laten zien in een **boxplot**. Een boxplot heeft dus vijf grenzen:

- Linkergrens met het laagste getal.
- Rechtergrens met het hoogste getal.
- Middelste grens  $Q_2$ , de mediaan.
- De tweede grens  $Q_1$  tussen de linkergrens en  $Q_2$ ; de mediaan van de eerste helft.
- De vierde grens  $Q_3$  tussen  $Q_2$  en de rechtergrens; de mediaan van de tweede helft.



Figuur 4

De interkwartielf afstand is het verschil tussen het eerste kwartiel ( $Q_1$ ) en het derde kwartiel ( $Q_3$ ), dus  $Q_3 - Q_1$ .

### Voorbeeld 1

Bekijk het steelbladdiagram van de cijfers in een klas. Het is tegelijk een klassenindeling (eerste klasse  $2,0 - < 3,0$ ) en een overzicht van alle cijfers (in de tweede klasse zit twee keer het cijfer 3,9). Je kunt er ook een boxplot van maken. Laat dat zien.

2	4								
3	9	9							
4	4	4							
5	0	0	1	5	5	9	9		
6	2	4	4	4	6	6	8	9	
7	1	3	7						
8	2	5	6	6	6	6			
9	5								

Figuur 5

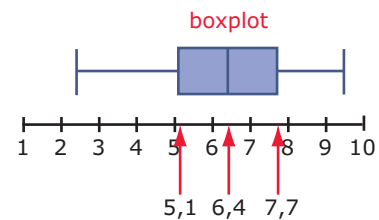
Antwoord

Bepaal eerst:

- De modus is 8,6 en de modale klasse is  $6,0 - < 7,0$ .
- De mediaan is 6,4 (het gemiddelde van de twee middelste cijfers), het eerste kwartiel is 5,1 en het derde kwartiel is 7,7.
- De spreidingsbreedte is  $9,5 - 2,4 = 7,1$  als je naar de werkelijke cijfers kijkt, of  $10,0 - 2,0 = 8,0$  als je naar de klassenindeling kijkt
- De kwartielf afstand is  $7,7 - 5,1 = 2,6$ . Het is de breedte van de box van de boxplot.

De boxplot verdeelt alle cijfers in vier delen met elk 25% van deze cijfers.

Bestaan de waarnemingen uit een oneven aantal waarden, dan wordt de mediaan van de hele set niet meegenomen om  $Q_1$  en  $Q_3$  te berekenen.



Figuur 6

### Opgave 3

Bekijk in **Voorbeeld 1** het steelbladdiagram en de boxplot. In de tabel zie je de cijfers gehaald voor een wiskundetoets van twee parallelklassen.

cijfers klas A					cijfers klas B				
6,7	6,4	4,9	3,8	4,0	4,0	6,2	4,9	3,9	5,9
5,6	5,8	6,8	8,2	4,7	7,3	4,7	6,7	7,6	9,4
3,4	8,5	4,1	6,9	7,3	8,3	5,7	7,2	8,7	7,1
6,1	7,5	6,7	6,2	3,4	7,0	6,5	7,4	5,0	4,8
7,9	4,5	8,3			7,7	6,5	4,9	8,8	6,3

Tabel 2

Maak van de cijfers van beide klassen een steelbladdiagram en bepaal de mediaan en de kwartielf afstand van beide klassen. Teken voor beide klassen een boxplot van de resultaten.

### Opgave 4

Welke uitspraak is waar voor de volgende waarnemingsgetallen?

58; 63; 51; 56; 86; 69; 55; 76; 74; 69; 45; 75; 55; 68; 68; 52; 70; 57; 65; 78; 65; 72; 83; 65; 79.

- A. De modus en mediaan zijn gelijk.
- B. De modus en het gemiddelde zijn gelijk.
- C. Het gemiddelde en de mediaan zijn gelijk.
- D. Geen van deze uitspraken is waar.

### Opgave 5

Welke uitspraken zijn waar voor de volgende waarnemingsgetallen?

58; 63; 51; 56; 86; 69; 55; 76; 74; 69; 45; 75; 55; 68; 68; 52; 70; 57; 65; 78; 65; 72; 83; 65; 79; 57; 63; 63; 72; 63.

- A. De modus is groter dan de mediaan.
- B. Het gemiddelde is groter dan de mediaan.
- C. De modus is kleiner dan het gemiddelde.

### Voorbeeld 2

Bekijk de tabel met leeftijd, lengte en gewicht van 36 mannen.

gegevens van 36 mannen											
nr	leeftijd (j)	lengte (cm)	gewicht (kg)	nr	leeftijd (j)	lengte (cm)	gewicht (kg)	nr	leeftijd (j)	lengte (cm)	gewicht (kg)
i	L	l	G	i	L	l	G	i	L	l	G
1	20	180	95	13	46	180	95	25	67	171	68
2	23	184,5	91	14	47	177,5	75	26	67	170	81
3	23	180,5	90	15	51	196	86	27	68	169	97
4	29	169	91	16	52	188,5	93	28	68	176	72
5	30	176	80	17	53	170	85	29	71	171	84
6	33	182	82	18	54	178,5	77	30	71	180,5	75
7	36	182,5	85	19	56	172,5	74	31	71	172	77
8	36	176,5	74	20	61	164	85	32	74	171	77
9	38	180	73	21	61	181	89	33	75	183	95
10	40	190,5	112	22	61	173	95	34	75	172,5	74
11	41	184	89	23	61	181	89	35	82	173	66
12	44	179	75	24	63	181,5	100	36	84	175	90

Tabel 3

Je wilt van alle drie de series waarnemingsgetallen zowel de drie centrummaten als de drie spreidingsmaten bepalen.

Gebruik je Excel, dan gebruik je de statistische functies om de centrummaten en de spreidingsmaten te (laten) berekenen. Zie het **Practicum**.

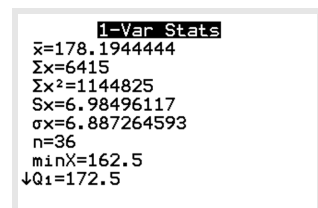
Werk je met de hand of met de grafische rekenmachine, dan maak je meteen geschikte frequentieverdelingen. Je kunt dan de centrummaten en de spreidingsmaten alleen nog schatten vanuit de klassenmiddens. Zie het **Practicum**.

#### Antwoord

Voor de modus en de mediaan zijn de sorteerfuncties in Excel erg handig. Verder kun je gemakkelijk optellen en kolommen maken met de waarden van een waarnemingsgetal maal zijn frequentie, enzovoort.

Werk je met de grafische rekenmachine, maak dan eerst een tabel van relatieve frequenties met bijbehorende klassenmiddens. Laat vervolgens de machine de centrum- en spreidingsmaten bepalen. Zo krijg je voor de lengte deze gegevens.

Bij de klassenindelingen zijn centrum- en de spreidingsmaten niet exact bepaald, omdat je door indelen in klassen de echte gegevens kwijt bent geraakt. Je krijgt zo schattingen van deze waarden.



Figuur 7

### Opgave 6

Bekijk in **Voorbeeld 2** de tabel met gegevens van 36 mannen en in het antwoord de tabel met centrummaten en spreidingsmaten.

- a Hoe wordt het gemiddelde berekend?
- b Hoe wordt de spreidingsbreedte (variatiebreedte) berekend?
- c Hoe wordt de kwartielf afstand berekend?
- d Ga na dat de modale leeftijd, de modale lengte en het modale gewicht correct zijn.
- e Bereken de kwartielen en teken de bijpassende boxplots.
- f Hoe wordt de standaardafwijking berekend?

### Opgave 7

Bekijk de tabel met gegevens van 36 vrouwen.

gegevens van 36 vrouwen											
nr	leeftijd (j)	lengte (cm)	gewicht (kg)	nr	leeftijd (j)	lengte (cm)	gewicht (kg)	nr	leeftijd (j)	lengte (cm)	gewicht (kg)
<i>i</i>	<i>L</i>	<i>l</i>	<i>G</i>	<i>i</i>	<i>L</i>	<i>l</i>	<i>G</i>	<i>i</i>	<i>L</i>	<i>l</i>	<i>G</i>
1	21	181	85	13	53	168,5	64	25	70	166	65
2	22	171	55	14	53	165	64	26	71	159	93
3	30	176	55	15	53	179	77	27	73	154	76
4	32	176,5	70	16	57	174	76	28	73	153	54
5	34	158,5	61	17	62	179	80	29	74	160	60
6	43	160	80	18	64	163,5	80	30	75	163,5	79
7	49	179	75	19	65	160	92	31	75	153	93
8	50	166	74	20	65	182	78	32	76	159	81
9	50	171	84	21	65	154	74	33	77	165	68
10	50	166,5	65	22	67	162	64	34	78	163	89
11	51	171	80	23	68	160	63	35	80	166,5	60
12	52	171	71	24	69	163	75,5	36	80	157	72

Tabel 4

- a Bepaal vanuit de basisgegevens de centrummaten en de kwartielen van de leeftijden, de lengtes en de gewichten van vrouwen.
- b Bepaal ook de bijbehorende spreidingsmaten en teken de boxplots.

### Opgave 8

In twee verzorgingstehuizen heeft men de keuze uit verschillende zitbreedtes voor rolstoelen. Van alle zitbreedtes zijn evenveel rolstoelen beschikbaar.

Verzorgingstehuis Alfa heeft rolstoelen met zitbreedtes van 35 cm, 38 cm, 41 cm, 45 cm en 50 cm. Verzorgingstehuis Omega heeft rolstoelen met zitbreedtes van 34 cm, 38 cm, 41 cm, 46 cm en 50 cm.

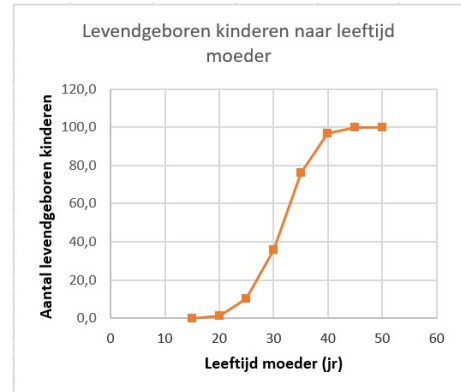
- a Bereken per verzorgingstehuis het gemiddelde en de mediaan van de zitbreedtes van de rolstoelen.
- b Geven deze centrummaten het verschil tussen de zitbreedtes van de rolstoelen van beide verzorgingstehuizen voldoende aan? Licht je antwoord toe.

### Voorbeeld 3

Je ziet een cumulatieve relatieve frequentiepolygoon bij deze klassenindeling van levendgeborenen naar leeftijd van de moeder.

Maak een bijpassende boxplot.

levendgeborenen naar leeftijd moeder 2003	
klasse	frequentie
15 – 19 jaar	2182
20 – 24 jaar	17383
25 – 29 jaar	49344
30 – 34 jaar	79001
35 – 39 jaar	39665
40 – 44 jaar	6225
45 – 49 jaar	217



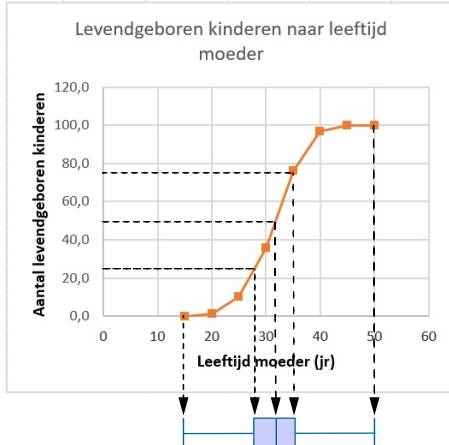
Figuur 8

Tabel 5

Antwoord

Lees bij 50% de mediaan af, bij 25% het eerste kwartiel en bij 75% het derde kwartiel. Het minimum en maximum zitten bij 0% en 100%. Je ziet:

- het eerste kwartiel  $Q_1 = 28$
- de mediaan  $Q_2 = 31,5$
- het derde kwartiel  $Q_3 = 35$



Figuur 9

### Opgave 9

Bestudeer hoe je bij een klassenindeling een boxplot maakt.

Bekijk vervolgens de tabel met gegevens van 36 mannen. Maak bij de klassenindeling van de kolom 'lengte' een cumulatieve relatieve frequentiepolygoon en een boxplot.

## Verwerken

### Opgave 10

Je hebt de waarnemingsgetallen 16,18,22,24,26,26,28,30 en 36.

- Teken een bijpassend boxplot.
- Doe dat nog eens als je bij alle getallen 4 optelt.
- En ook als je van alle getallen 40 aftrekt.
- Doe het nog eens als je alle getallen door 2 deelt.
- Welk resultaat krijg je als je alle getallen met 3 vermenigvuldigt?
- Beschrijf wat er met de boxplot gebeurt als bij alle waarnemingsgetallen een getal wordt opgeteld of ervan afgetrokken wordt.
- Beschrijf wat er met de boxplot gebeurt als alle waarnemingsgetallen met een getal worden vermenigvuldigd of door een getal worden gedeeld.

### Opgave 11

Voor een practicum biologie worden regenwormen gevangen. De lengte van die regenwormen vind je in de tabel.

- Kijk naar de manier waarop de klassen zijn gemaakt. Hoe nauwkeurig zijn de regenwormen gemeten? Bij welke klasse hoort een regenworm die 3,0 cm lang is?
- Welke klasse is de modale klasse?
- Teken een histogram van de cumulatieve relatieve frequenties. Teken in dezelfde figuur de cumulatieve frequentiepolygoon.
- In welke klasse zit de mediaan? Kun je precies zeggen hoe groot die mediaan is? Schat de mediaan met behulp van de cumulatieve frequentiepolygoon.
- Bereken het gemiddelde en de standaardafwijking.

lengte regenworm (cm)	aantal
0,0– < 3,0	4
3,0– < 6,0	8
6,0– < 9,0	17
9,0– < 12,0	22
12,0– < 15,0	23
15,0– < 18,0	17
18,0– < 21,0	6
21,0– < 24,0	2
24,0– < 27,0	1

Tabel 6

### Opgave 12

Een supermarkt laat onderzoek verrichten naar de besteding per klant en naar de hoeveelheid tijd die een klant aan de kassa nodig heeft om af te rekenen. Er worden op verschillende tijdstippen tellingen gehouden. Je ziet de resultaten.

kassatijd (min)	aantal klanten	kassabon (€)	aantal klanten
0– < 1	22	0– < 50	24
1– < 2	75	50– < 100	61
2– < 3	58	100– < 150	75
3– < 4	25	150– < 200	25
4– < 5	16	200– < 250	12
5– < 6	4	250– < 300	2
		300– < 350	1

Tabel 7

- Bepaal bij beide tabellen de modus, de mediaan, het eerste en het derde kwartiel en het gemiddelde.
- Hoe groot is de standaardafwijking bij beide verdelingen?
- Teken bij beide tabellen een boxplot.



De supermarkt heeft een weekomzet van € 150000,00. Een caissière mag 38 uur per week werken.

- d** Hoeveel caissières moet de supermarkt in dienst nemen als er vanwege de wisselende winkeldrukke een overcapaciteit van 25% wordt aangehouden?

### Opgave 13

Op elk uur van een dag is de temperatuur bepaald. De uren van middernacht tot 12 uur 's middags worden aangegeven met am (het Latijnse 'ante meridiem' (am) betekent 'voor het middaguur'), en de uren van 12 uur 's middags tot middernacht met pm ('post meridiem' (pm)).

uur	1	2	3	4	5	6	7	8	9	10	11	12
am	16,1	15,8	15,4	15,1	14,8	15,0	16,1	17,4	18,5	19,4	20,3	21,1
pm	21,9	22,6	22,7	22,5	21,9	21,0	19,8	18,8	18,1	17,5	17,0	16,6

Tabel 8

- a** Verwerk deze gegevens in een dubbel steelbladdiagram.
- b** Maak boxplots van elk dagdeel afzonderlijk en van de totale dag.
- c** Bereken voor beide dagdelen afzonderlijk het gemiddelde en de standaardafwijking.
- d** Bereken het gemiddelde en de standaardafwijking van alle metingen van die dag.
- e** Geef een verklaring voor de verschillen die je vindt. (Dit is bivariate statistiek. Je bekijkt twee variabelen en hun eventuele samenhang/verschil.)

### Opgave 14

In een nieuw te bouwen ziekenhuis moeten bedden worden aangeschaft. De facilitaire dienst vraagt zich af welke lengte de bedden moeten krijgen. Hoe langer de bedden, hoe hoger de kosten. In het oude ziekenhuis hebben ze het laatste jaar van 278 patiënten gegevens verzameld. Je vindt ze in het bestand [Patiëntengegevens](#).

- a** Bereken de gemiddelde lengte van de patiënten. Bereken ook de gemiddelde lengte van de vrouwelijke en de mannelijke patiënten apart.
- b** Men kan natuurlijk alle bedden zo lang maken als de langste patiënt. Hoe lang worden de bedden dan? Noem een bezwaar tegen dit idee.

Handiger is misschien de lengte van het bed zo te kiezen dat 50% van de patiënten erin past. Voor langere patiënten neem je dan een bed met een lengte van de langste patiënt.

- c** Hoe lang moet het bed dan worden als 50% van de patiënten erin past?  
Het hoofd van de facilitaire dienst denkt dat het goedkoper is om verschil te maken in mannen- en vrouwenbedden, wat de lengte betreft.
- d** Als we hiervan uitgaan en de voorwaarde na b, hoe lang worden dan een 'mannenbed' en een 'vrouwenbed'?

### Opgave 15

Het gemiddelde van de getallen 23, 6, 15, 31, 7 en  $c$  is gelijk aan 16. Bereken  $c$ .

### Opgave 16

In een eetcafé komen op een dag tachtig mensen eten. Ze besteden gemiddeld € 22,00. Studenten betalen € 15,00 voor het studentenmenu, dagjesmensen betalen € 25,00 voor het toeristenmenu. Hoeveel studenten waren er die dag?

## Toepassen

In 1947 zijn bij 5001 vrouwelijke klanten van de Bijenkorf vijftien lichaamsmaten opgemeten. Vervolgens is gekeken welke van deze maten het meest bruikbaar zijn om een **maatsysteem voor kleding** op te baseren. Bekijk een deel van de uitkomst van het onderzoek in het bestand [Statistiek Bijenkorf 1947](#).

Je kunt met Excel centrum en spreiding van de gegevens berekenen en boxplots maken.

Soms is er sprake van een **uitschieter**, een waarde die wel erg veel afwijkt.

In een boxplot is een uitschieter een waarde die meer dan 1,5 keer de interkwartielafstand onder het eerste kwartiel of boven het derde kwartiel zit.

### Opgave 17

Bekijk bij [Toepassen](#) de gegevens over lengte en gewicht van de 5001 gemeten vrouwen.

- Bepaal de modale lengte en bereken de gemiddelde lengte.
- Bepaal de mediaan en de kwartielen van de lengte en teken een bijpassende boxplot.
- Zijn er bij deze lengtes uitschieters? Licht je antwoord toe.
- Verdeel de lengtes in klassen van 5 cm, te beginnen bij 135– < 140. Maak bij de nieuwe frequentieverdeling een relatieve cumulatieve frequentiepolygoon. Welke vorm heeft de polygoon?
- Bepaal opnieuw de gemiddelde lengte, de mediaan en de kwartielen. Wijken de resultaten veel af van de antwoorden bij a en b?

### Opgave 18

In de dataset in [Toepassen](#) vind je ook gegevens over de *mouwlengte* en de *kniehoogte* van de 5001 vrouwen.

- Bepaal modus, mediaan en het gemiddelde van de mouwlengtes  $m_i$ .
- Bepaal de spreidingsbreedte, de interkwartielafstand en de standaardafwijking van de mouwlengtes  $m_i$ .
- Er is één nogal afwijkende mouwlengte. Is hier sprake van een uitschieter?

## Testen

### Opgave 19

Op een feestje zijn acht personen aanwezig. Je ziet een tabel met gegevens over de feestgangers.

naam	leeftijd in jaar	lengte in cm	favoriete drankje	gewicht in kg	zakgeld in € per week	vervoermiddel
Jan	15	187	cola	71	€ 15	fiets
Leo	16	178	colal	69	€ 20	fiets
Elske	15	172	sinas	66	€ 17	bus
Daphne	17	171	cola	67	€ 20	scooter
Mart	19	195	cassis	76	€ 19	auto
Durk	57	180	bier	74	€ 50	auto
Henk	17	187	cola	75	€ 18	fiets
Leo	15	179	cola	70	€ 18	brommer

Figuur 10

- Welke centrummaat zou je gebruiken om voor elke kolom de feestgangers te typeren?
- Welke spreidingsmaat zou je zo mogelijk gebruiken om voor elke kolom de feestgangers te typeren?
- Hoe zou je de doorsneefeestganger omschrijven?

### Opgave 20

Een groep leerlingen wordt tijdens de zwemles gevraagd om zo lang mogelijk hun adem in te houden onder water. Je ziet hier de bijbehorende tijden (in seconden).

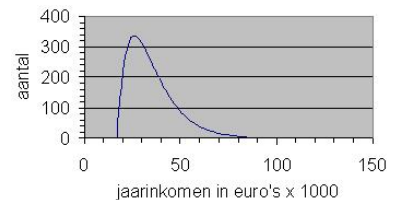
17; 39; 17; 21; 26; 21; 31; 17; 37; 43  
 36; 17; 15; 29; 21; 31; 35; 23; 18; 17  
 26; 21; 28; 23; 22; 16; 37; 33; 27

- Bepaal de drie centrummaten van deze gegevens.
- Bepaal de spreidingsbreedte en de standaardafwijking.
- Teken een boxplot bij deze gegevens.
- Maak een klassenindeling met de eerste klasse  $15 - < 20$ .
- Bereken het gemiddelde en de standaarddeviatie bij deze klassenindeling.
- Teken een bijpassende cumulatieve relatieve frequentiepolygoon. Bepaal daarmee de mediaan bij deze klassenindeling.

### Opgave 21

In de grafiek vind je de jaarinkomens van de werknemers van een grote fabriek.

- Wat verdient de doorsneewerknemer van deze fabriek? Welke centrummaat heb je gekozen en waarom?
- Welke centrummaat is groter, de mediaan of de modus? Leg uit waar je dat aan kunt zien.



Figuur 11

## Practicum

Met de volgende practica kun je **de statistische functies van de grafische rekenmachine** door-nemen. Onder andere staat er in hoe je centrummaten en spreidingsmaten kunt berekenen.

- [Statistiek met de TI84](#)
- [Statistiek met de TIinspire](#)
- [Statistiek met de Casio fx-CG50](#)
- [Statistiek met de HPprime](#)
- [Statistiek en de NumWorks](#)


De **statistische functies van Excel** vind je in het volgende practicum. Er staat meer in dan op dit moment nodig is, maar onder andere kun je er nog eens in vinden hoe je diagrammen maakt en centrummaten en spreidingsmaten kunt laten berekenen door Excel.

- [Data presenteren](#)



© 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 Foliostaat 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.

---