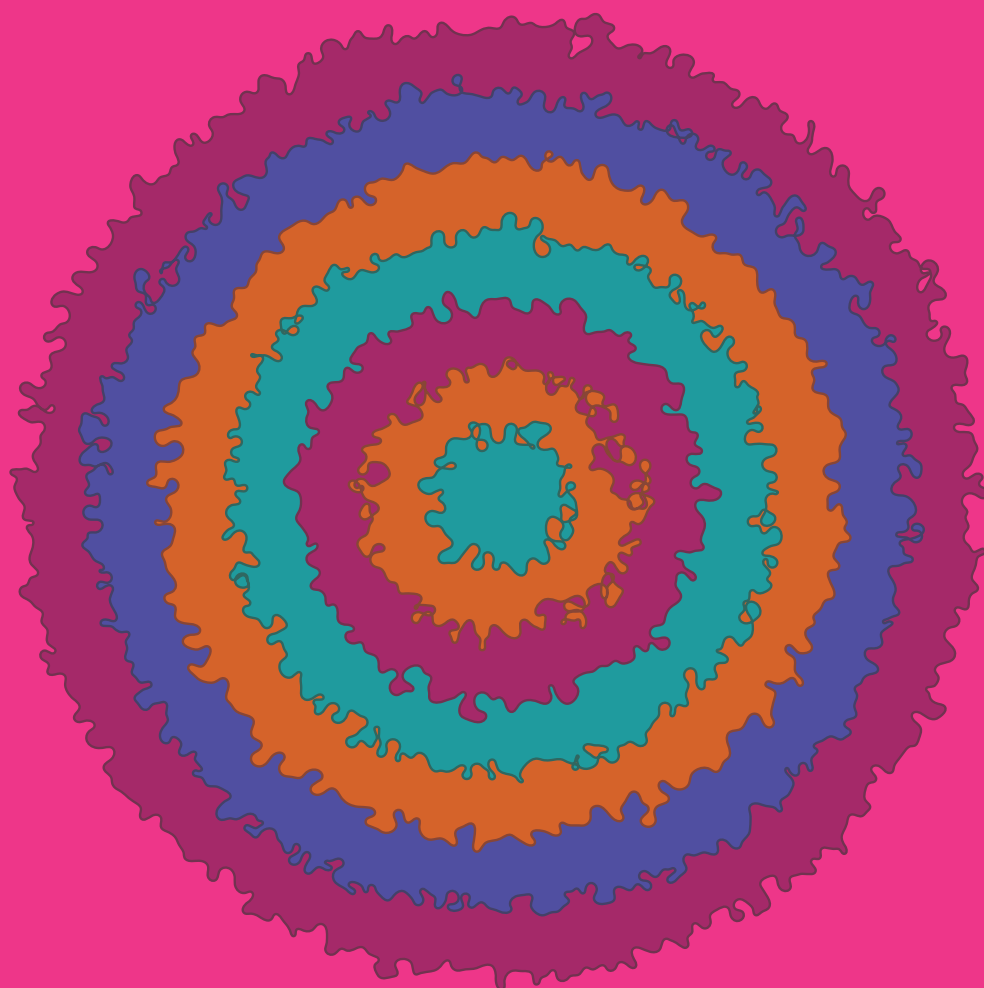


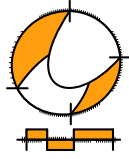
Wiskunde / PGA

2 HAVO / VWO / docentmateriaal

Statistiek

ConTeXt College





© 2024

Het auteursrecht op dit lesmateriaal berust bij Stichting Math4All. Math4All is derhalve de rechthebbende zoals bedoeld in de hieronder vermelde creative commons licentie.

Het lesmateriaal is met zorg samengesteld en getest. Stichting Math4All aanvaardt geen enkele aansprakelijkheid voor onjuistheden en/of onvolledigheden in de module. Ook aanvaardt Math4All geen enkele aansprakelijkheid voor enige schade, voortkomend uit (het gebruik van) dit lesmateriaal

Voor deze module geldt een Creative Commons Naamsvermelding Niet Commercieel 3.0 Nederland Licentie. (zie <http://creativecommons.org/licenses/by/3.0>).

Dit lesmateriaal is open, gratis en vrij toegankelijk lesmateriaal afkomstig van Stichting Math4All en is speciaal ontwikkeld voor het vak wiskunde in het voortgezet onderwijs. Het lesmateriaal op de website www.math4all.nl is afgestemd op kerndoelen wiskunde, tussendoelen wiskunde en eindtermen voor de vakken wiskunde A, B en C. Dit lesmateriaal is mediumneutraal ontwikkeld en op diverse manieren te bekijken en te gebruiken. Voor informatie en vragen kunt u contact opnemen via info@math4all.nl. Ook houden we ons altijd aanbevolen voor suggesties, verbeteringen en/of aanvullingen.

Voorwoord

Het lesmateriaal in dit katern is gebaseerd op het materiaal dat je kunt vinden op de Math4All website www.math4all.nl. In de tekst staan dan ook regelmatig verwijzingen naar die website. Waar je precies moet zijn op die website kun je zien in de kopregel van iedere pagina.

Ieder hoofdstuk bestaat uit een aantal paragrafen en wordt steeds afgesloten met een paragraaf *Totaalbeeld* waar de leerstof wordt samengevat en/of herhaald.

PGA

PGA staat voor 'probleemgestuurde aanpak'. Je begeleidt dan als docent de leerlingen die in kleine groepjes aan wiskundige problemen werken en op die manier een eigen theoretisch kader opstellen. Dit gebeurt voornamelijk op de wijze die wordt beschreven in het boek *Building Thinking Classrooms in Mathematics* van Peter Liljedahl. Dit boek is ook in het Nederlands beschikbaar. Het is verstandig om dit boek vooraf door te werken, maar je kunt ook beginnen met deze **beknopte handleiding**.

De PGA wordt ondersteund door verwerkings- en toepassingsopgaven waarmee de leerling kan nagaan of de stof wordt beheersd. Deze opgaven worden op drie niveaus aangeboden. De niveau aanduiding staat in de marge naast de opgave.

- ★ het basale niveau, dat iedereen zou moeten behalen
- ★ ★ een iets pittiger niveau, waarin iets meer uitdaging zit en die de leerling alleen hoeft te maken als er genoeg tijd voor is
- ★ ★ ★ een bijzondere toepassing of een echt pittige opgave die een leerling alleen maakt als de rest veel te gemakkelijk was

In de bijlage staat een "**Leerdoelentabel**" waarin staat aangegeven door welke opgave het specifieke leerdoel wordt afgedekt en op welk niveau dit gebeurt. Als je deze tabel aan de leerlingen uitreikt, kunnen ze hun eigen vorderingen bijhouden.

Opgaven uit de samenvattende paragraaf *Totaalbeeld* worden voorafgegaan door een T.

1

Statistiek

1.1	Centrummaten	6
1.2	Spreidingsmaten	14
1.3	Klassenindeling	23
1.4	Schattingen	31
1.5	Statistische uitspraken	38
1.6	Totaalbeeld	44

1.1 Centrummaten

Inleiding

Als je beschikt over een hele serie gegevens, zoals alle rapportcijfers van alle leerlingen in leerjaar 2 van alle vakken, dan heb je een enorme brij aan getallen. Hoe krijg je daar enig overzicht over?

Zoals je weet kunnen frequentietabellen en diagrammen helpen. Maar soms is een enkel getal genoeg...

B1H		ne		en		fa		ak		gs	
leerling	geslacht	RE	RE	RE	RE	RE	RE	RE	RE	RE	RE
1	v	6,7	7	4,4	4	5,6	6	6,6	7	6,8	7
2	v	5,6	6	5,3	5	6,1	6	7,1	7	6,8	7
3	m	8,1	8	6,7	7	5,8	6	7,2	7	7,6	8
4	m	8,5	9	5,1	5	6,1	6	6,1	6	6,1	6
5	m	4,9	5	9,7	10	6,6	7	8,0	8	7,5	8
6	v	6,2	6	9,4	9	7,2	7	6,6	7	7,8	8

Figuur 1.1

Je leert in dit onderwerp

- van een hoeveelheid gegevens de modus, de mediaan en het gemiddelde berekenen;
- modus, mediaan en gemiddelde interpreteren als centrummaten van die gegevens.

Voorkennis

- (gewogen) gemiddelde van een hoeveelheid gegevens berekenen;
- werken met (relatieve) frequenties en (relatieve) frequentietabellen;
- werken met kruistabellen.

Voor de docent

Bij het onderdeel 'Centrummaten' gaat het er om bij een serie gegevens, een dataset, de drie verschillende maten voor het centrum van die gegevens te leren kennen en te berekenen. Ook wordt een eerste stap gezet naar het interpreteren van de verschillende centrummaten.

Het is wellicht mooier om bij de eerste twee opdrachten een set eindcijfers van enkele klassen (van een jaar geleden, of nog ouder, geanonimiseerd) van de eigen school te gebruiken.

Gewenste materialen:

- Maak vooraf een kopie van het werkblad bij de eerste, de tweede en de derde opdracht.
- Schrijfmateriaal voor op de verticale werkvlakken en eventueel plakband om de informatiebladen eraan op te hangen.

Opdracht 1.1

Je ziet hier de rapportcijfers voor het vak Engels in B1H.

4	5	7	5	10	9	8	4	6	6	7	7	8	3	7
6	6	7	7	9	7	6	7	6	6	8	7	7	8	

Tabel 1.1

Het cijfer dat het meest voor komt, heet het 'modale cijfer', ook wel de 'modus' genoemd. Het middelste cijfer (als je ze van klein naar groot op volgorde zet) heet de 'mediaan'. En verder kun je nog het 'gemiddelde' van een serie cijfers berekenen.

Maak van deze serie cijfers een frequentietabel en bereken daarmee het modale cijfer, de mediaan en het gemiddelde cijfer. Leg uit, welk van deze drie getallen de beste 'centrummaat' van de verdeling van de cijfers is.

— **Toelichting** —

Geef de opdracht mondeling. Kopieer de gegevens vooraf vanaf het **Werkblad** en deel dit uit.

Wellicht moeten begrippen als gemiddelde en frequentietabel (bekend uit het onderwerp ‘Diagrammen’) nog even in herinnering worden gebracht. Mogelijke hulpvragen: “Wat is een frequentietabel ook alweer?”, “Hoe bepaal je de frequentie van een cijfer?” en/of “Hoe bereken je ook alweer een gemiddelde vanuit een frequentietabel?”.

Benoem na afloop het begrip ‘centrummaat’ en zet de drie centrummaten op een rijtje. Bespreek ook het afronden van het gemiddelde: waarom zou je daar liever met één decimaal werken? Bekijk wat de verschillende groepen hebben gezegd over het nut van de verschillende centrummaten.

— **Uitwerking** —

Frequentietabel:

cijfer	frequentie
3	1
4	2
5	2
6	7
7	10
8	4
9	2
10	1
totaal	29

Tabel 1.2

Modale cijfer is 7.

Mediaan is 7.

Gemiddelde is $\frac{1 \cdot 3 + 2 \cdot 4 + 2 \cdot 5 + 7 \cdot 6 + 10 \cdot 7 + 4 \cdot 8 + 2 \cdot 9 + 1 \cdot 10}{29} \approx 6,7$.

Het gemiddelde, de mediaan en het modale cijfer geven alle drie wel een beeld, je wilt ze het liefst allemaal weten.

Opdracht 1.2

Je ziet hier de frequentietabellen van de klassen B2A en B2C van hun rapportcijfers voor het vak wiskunde. Vergelijk de modus, de mediaan en het gemiddelde cijfer van beide klassen. Trek conclusies over welke klas gemiddeld beter scoort, in welke klas zitten de beste leerlingen, waar vallen de meeste onvoldoendes, etc.

	klas B2A	klas B2C
cijfer	frequentie	frequentie
4	1	0
5	4	4
6	9	8
7	11	6
8	3	4
9	1	2
totaal	29	24

Tabel 1.3

— **Toelichting** —

Geef de opdracht mondeling. Kopieer vooraf het **Werkblad** en deel dit uit.

Hierbij gaat het om het vergelijken van twee datasets op basis van de centrummaten. Omdat B2C een even aantal leerlingen kent, moet hier de mediaan worden berekend door het gemiddelde van de middelste twee getallen te nemen. Daarbij is wellicht enige hulp nodig.

— **Uitwerking** —

Klas B2A:

- de modus is 7, want 7 komt het vaakst voor bij klas B2A.
- de mediaan is 7, want dat is het 15de getal als je ze van klein naar groot opschrijft.
- het gemiddelde is ongeveer 6,5, want $\frac{1 \cdot 4 + 4 \cdot 5 + 9 \cdot 6 + 11 \cdot 7 + 3 \cdot 8 + 1 \cdot 9}{29} \approx 6,5$.

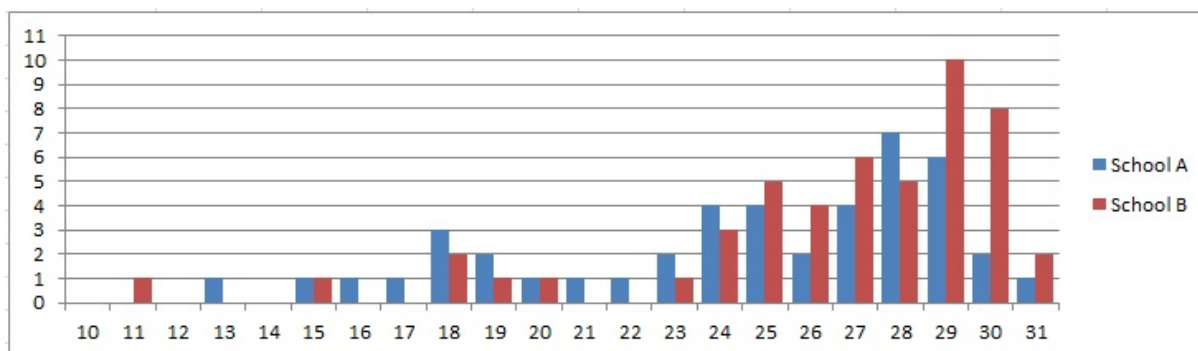
Klas B2C:

- de modus is 6, want 6 komt het vaakst voor bij klas B2C.
- de mediaan is 6,5, want dat is het gemiddelde van het 12e en 13e cijfer als je ze van klein naar groot opschrijft.
- het gemiddelde is ongeveer 6,7, want $\frac{4 \cdot 5 + 8 \cdot 6 + 6 \cdot 7 + 4 \cdot 8 + 2 \cdot 9}{24} \approx 6,7$.

Klas B2C scoort gemiddeld hoger dan B2A, maar het meest voorkomende cijfer zit bij B2C onder het gemiddelde. Er zitten in klas B2C dus veel leerlingen die magertjes scoren (de helft scoort 5 of 6), maar ook naar verhouding veel leerlingen die heel goed scoren.

Opdracht 1.3

Op twee scholen voor voortgezet onderwijs zijn de aantallen leerlingen per klas geteld. Je ziet in dit staafdiagram het resultaat voor school A en school B.



Figuur 1.2

Bepaal de modale en de gemiddelde klassengrootte per school en gebruik deze gegevens om te bepalen welke school in het algemeen grotere klassen heeft.

— **Toelichting** —

Geef de opdracht mondeling. Kopieer het diagram vooraf vanaf het **Werkblad** en deel dit uit.

Deze opdracht gaat om het bepalen van de centrummaten in een dubbel staafdiagram. Mogelijke hulpvraag om het diagram te begrijpen: “Hoeveel klassen met 25 leerlingen heeft school A? En school B”.

— **Uitwerking** —

School A:

- de modale klassengrootte: 28, want de blauwe staven horen bij school A en de hoogste blauwe staaf hoort bij een klassengrootte van 28.
- de gemiddelde klassengrootte: ongeveer 24,5.

School B:

- de modale klassengrootte: 29.
- de gemiddelde klassengrootte: ongeveer 26,4.



De modus is bij school B net iets groter dan bij school A, bovendien heeft school B een grotere gemiddelde klassengrootte. Op school B zijn de klassen in het algemeen dus groter (ondanks die éne uitschieter van een klasje van maar 11 leerlingen).

Opdracht 1.4

Bekijk wat iedereen heeft gemaakt en heeft bedacht over het bepalen van 'centrummaten' bij een verzameling data (gegevens) en welke centrummaten er zijn.

Maak een eigen overzicht van wat je hebt geleerd.

Toelichting

Loop samen met de leerlingen alle bedenksels na. Bevraag leerlingen of ze elkaars gedachtenspingsels kunnen toelichten. Samen zouden jullie naar een overzicht van de theorie moeten komen. Ieder schrijft het voor zichzelf op.

Uitwerking

Het theorieblok geeft het gewenste overzicht.

Theorie

Om te onthouden

Er zijn drie getallen die een hoeveelheid gegevens (waarnemingen) kunnen samenvatten:

- de **modus** (de modale waarde) is de waarneming die het vaakst voorkomt en dus de grootste frequentie heeft. Er is geen modus als twee waarnemingen het meest voorkomen.
- de **mediaan** is het midden van alle waarnemingen als die op volgorde staan.
- het **gemiddelde** van de waarnemingen vind je door alle waarden bij elkaar op te tellen en dat getal te delen door het totaal aantal cijfers. Daarbij moet je rekening houden met de frequenties (de wegingen) van de waarnemingen.

De modus, de mediaan en het gemiddelde zijn **centrummaten**. Deze drie getallen kunnen verschillend zijn.

Het is niet zo dat je modus, mediaan en gemiddelde altijd kunt bepalen. Je moet voor de mediaan en het gemiddelde altijd getallen als waarneming hebben.

Verwerken

★ Opgave 1.1

Dit zijn de rapportcijfers die in een bepaalde klas voor economie zijn behaald:

7	4	6	6	5	5	7	6	7	9	10	6	8	7
8	6	5	7	5	8	3	7	8	6	6	10	5	7

Tabel 1.4

- Maak een frequentietabel van de rapportcijfers voor economie.
- Bepaal de modus.
- Bepaal de mediaan.

Twee leerlingen die tijdens de toets ziek waren maken een inhaaltoets. Ze halen allebei een zeven.

- Bereken het nieuwe gemiddelde, de nieuwe modus en de nieuwe mediaan.

★ Opgave 1.2

Een centrummaat moet een goede indruk geven van de waarnemingen die je bestudeert. Kies in de volgende gevallen een geschikte centrummaat, en bereken deze maat.

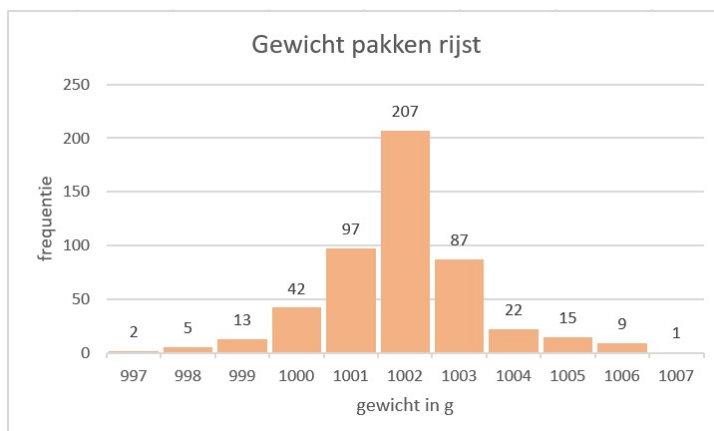
- Erik haalde voor zijn wiskundetoetsen de cijfers 8, 5, 6, 7, 10, 3, 7, 7 en 8.
- De directeur van een onderneming krijgt € 8000 per maand. De vier adjunctdirecteuren ontvangen maandelijks € 3000. De overige dertig werknemers verdienen elk € 1900 per maand.
- Van de docenten van een scholengemeenschap in Overdal komen er 16 te voet naar school, 20 komen met de fiets, 25 met de auto en 37 met het openbaar vervoer.
- De resultaten van een meting van de maximumtemperaturen in een week in juli in graden Celsius waren:

dag	ma	di	wo	do	vr	za	zo
temperatuur in °C	22	23	24	23	23	23	23

Tabel 1.5

★ Opgave 1.3

In een supermarkt worden pakken rijst verkocht met een gewicht van 1 kg. Elke werkdag wordt een vast aantal pakken gecontroleerd op afwijkend gewicht. In het staafdiagram zijn de resultaten verwerkt van 20 werkdagen.



Figuur 1.3

- Bepaal de modus en de mediaan.
- Hoeveel pakken werden er elke werkdag gecontroleerd?

- c Hoeveel procent van de gecontroleerde pakken rijst had een te laag gewicht?
- d Bereken het gemiddelde aantal pakken rijst per dag waarvan het gewicht te laag is.
- e In totaal zijn in het afgelopen jaar 8500 pakken rijst verkocht. Geef een schatting van het aantal pakken dat minstens het juiste gewicht had.

★ **Opgave 1.4**

Iemand heeft van een aantal doosjes lucifers geteld hoeveel lucifers er in zitten. De resultaten staan in de tabel hiernaast.

- a Van hoeveel doosjes lucifers is het aantal lucifers geteld?
- b Hoeveel lucifers zijn er in totaal geteld?
- c Hoeveel lucifers zitten er gemiddeld in een doosje? Geef je antwoord in één decimaal nauwkeurig.
- d Waarom wist je voordat je de berekening bij c uitvoerde dat het gemiddelde lager moest zijn dan de modus 36 ?

aantal lucifers	frequentie
31	1
32	3
33	4
34	5
35	7
36	8
37	11
38	6

Tabel 1.6

★ **Opgave 1.5**

Een loterij heeft op een dag aan prijzen gemiddeld € 20000 uitbetaald. De mediaan van die prijzen blijkt € 1500 te zijn.

Geef een verklaring voor het grote verschil tussen die twee centrummaten.

★★ **Opgave 1.6**

Een school kent drie rapportperiodes die alle drie even zwaar tellen. Voor wiskunde zijn er de eerste periode drie toetsen gegeven die elk 1 keer, 1 keer en 3 keer meetellen. In de tweede periode zijn er in dat vak twee toetsen gegeven die 2 keer en 3 keer meetellen en in de derde en laatste periode zijn er twee toetsen die 1 keer meetellen en een proefwerk dat drie keer meetelt. Je staat voor wiskunde een 7,2 en alleen het laatste proefwerk moet nog worden gemaakt. De cijfers voor alle toetsen worden in één decimaal nauwkeurig bepaald.

Kun je nog een 8 halen als eindcijfer? Laat met een berekening zien welk cijfer je dan voor het laatste proefwerk moet halen.

★★★ **Opgave 1.7**

In klas V2A zitten 20 leerlingen. Voor een toets Engels scoorde de klas gemiddeld 6,6. Ook klas V2B maakte die toets, maar hier was het gemiddelde cijfer 8,1. Het gemiddelde cijfer van de twee klassen samen was 7,5.

- a Leg uit waarom het gemiddelde niet gelijk is aan $\frac{6,6+8,1}{2} = 7,35$.
- b Hoeveel leerlingen zitten er in klas V2B?
- c De jongens in klas V2A scoorden gemiddeld 6,4 en de meisjes 6,9. Hoeveel jongens zitten er in klas V2A?

Toepassen

★ ★ Opgave 1.8: Centrummaten met een spreadsheet

Een spreadsheet (rekenbladprogramma) kan je helpen bij het maken van frequentietabellen en het berekenen van gemiddelden. Bekijk het [Practicum](#). Daarin werk je met het een bestand dat is gemaakt voor de spreadsheet waarmee je werkt.

Denk er wel om dat je op een pc of een tablet dit bestand eerst moet downloaden. Er wordt van uit gegaan dat je wel eens eerder met een spreadsheet hebt gewerkt en al weet hoe je met cellen werkt, met formules in cellen werkt, absoluut en relatief kopiëren begrijpt, etc.

- a Maak zelf een frequentietabel van de gehele eindcijfers voor wiskunde van deze groep leerlingen in 4 vwo.
- b Zoek uit hoe je van een rij gegevens het gemiddelde, de modus en de mediaan door Excel kunt laten berekenen. Het gaat met behulp van formules. Zoek een geschikte plek in het Excel-bestand om deze centrummaten neer te zetten.
- c Bereken het gemiddelde voor wiskunde nog eens met behulp van de frequentietabel.
- d Maak vervolgens afzonderlijke groepen van de jongens en de meisjes door te sorteren. Bereken de centrummaten voor deze afzonderlijke groepen en vergelijk ze met elkaar. Maak ook frequentietabellen en staafdiagrammen van de afzonderlijke groepen.

★ ★ Opgave 1.9: Geboortemaanden

Werk met hetzelfde Excel-bestand als in de voorgaande opgave. Zet de gegevens over de geboortemaanden op een afzonderlijk werkblad.

- a Maak een frequentietabel van de geboortemaanden van deze groep leerlingen in 4 vwo.
- b Welke centrummaat is hier zinvol?
- c Probeer een conclusie te trekken uit de frequentietabel. Of maak er eerst een staafdiagram van en probeer dan een conclusie te trekken.

Practicum: Statistiek met een spreadsheet

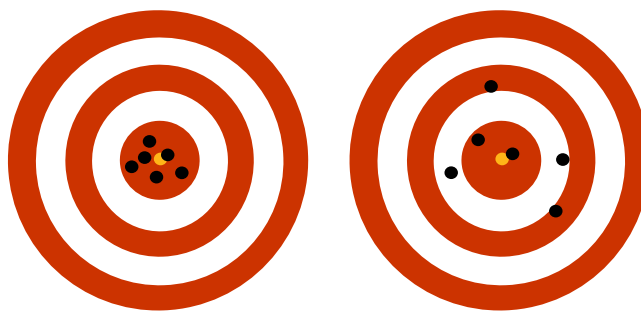
Een **spreadsheet of rekenbladprogramma** kan centrummaten voor je berekenen vanuit de data. In de volgende practica leer je werken met statistiek in een spreadsheet. Download wel eerst het bijbehorende bestand "Gegevens 154 leerlingen".

- [Data presenteren met behulp van MS-Excel](#)
- [Data presenteren met behulp van Open Office Calc](#)
- [Data presenteren met behulp van Google spreadsheets](#)

1.2 Spreidingsmaten

Inleiding

Ook al hebben twee groepen gegevens dezelfde centrummaat, dan nog kunnen er grote verschillen zijn. De spreiding kan erg verschillend zijn. Je ziet dat hier met de schoten op deze roos. Van de tweede serie is de spreiding groter.



Figuur 2.1

Je leert in dit onderwerp

- de spreidingsbreedte en de kwartielafstand van een serie gegevens berekenen en interpreteren als spreidingsmaten;
- een serie gegevens samenvatten in een boxplot.

Voorkennis

- (gewogen) gemiddelde, modus en mediaan van een hoeveelheid gegevens berekenen;
- werken met (relatieve) frequenties en (relatieve) frequentietabellen;
- werken met kruistabellen.

Voor de docent

Bij het onderdeel 'Spreidingsmaten' gaat het er om bij een serie gegevens, een dataset, twee verschillende maten voor de spreiding van die gegevens te leren kennen en te berekenen. Ook wordt die spreiding weergegeven in een boxplot en worden datasets vergeleken met behulp van boxplots. Het is wellicht mooier om bij de eerste twee opdrachten een set eindcijfers van enkele klassen (van een jaar geleden, of nog ouder, geanonimiseerd) van de eigen school te gebruiken.

Gewenste materialen:

- Maak vooraf een kopie van het werkblad bij de eerste, de tweede en de derde opdracht.
- Schrijfmateriaal voor op de verticale werkvlakken en eventueel plakband om de informatiebladen eraan op te hangen.

Opdracht 2.1

Je ziet hier de rapportcijfers voor het vak Engels in B1H.

cijfer	frequentie
3	1
4	2
5	2
6	7
7	10
8	4
9	2
10	1
totaal	29

Tabel 2.1

Het middelste cijfer is de ‘mediaan’. Maar je kunt deze cijfers ook in vier groepen verdelen. Het ‘eerste kwartiel’ is de mediaan van de eerste helft, het ‘tweede kwartiel’ is de mediaan en het ‘derde kwartiel’ is het midden van de tweede helft. Verder hebben deze cijfers een ‘minimum’ (laagste waarde) en een ‘maximum’ (hoogste waarde). Al deze getallen kun je op een getallenlijn plaatsen en zo de spreiding van de getallen weergeven.

Maak zo'n figuur en zeg iets over de spreiding van deze gegevens.

Toelichting

Geef de opdracht mondeling. Kopieer de gegevens vooraf vanaf het **Werkblad** en deel dit uit.

Het kan zijn dat verschillende groepen op verschillende manieren de kwartielen berekenen: tel je bijvoorbeeld bij de eerste helft de mediaan nog weer mee? En tel je die dan ook bij de tweede helft mee? De afspraak dat de mediaan niet meetelt bij de helften kun je zelf laten vallen, je kunt dit ook achteraf bespreken.

Bespreek na afloop het begrip ‘boxplot’ en de gewenste vorm van een boxplot. Benoem ook de verschillende lengtes van de kwarten en wat je daaraan kunt zien en laat de begrippen ‘spreidingsbreedte’ en ‘(inter)kwartielafstand’ vallen.

Uitwerking

Het minimum is 3.

Het eerste kwartiel is 6 (het gemiddelde van het 7de en het 8ste getal).

De mediaan is 7.

Het derde kwartiel is 7,5 (het gemiddelde van het 21ste en het 22ste getal).

Het maximum is 10.

Een fraaie tekening (liefst een echte boxplot) van deze waarden op/boven de getallenlijn.

Uitspraken kunnen zijn: “De meeste cijfers liggen dicht bij de mediaan.”, “De helft van de cijfers ligt tussen 6 en 7,5.”, “Een kwart van de cijfers ligt boven de 7,5.”, etc.

Opdracht 2.2

Je ziet hier een kruistabel van de cijfers voor de vakken wiskunde en scheikunde in B2D.

Vergelijk de resultaten van deze twee vakken met behulp van twee boxplots.

		wi							
		B2d	4	5	6	7	8	9	10
sc	4		1						
	5			2	3	1			
	6	1	4	2	1	1			
	7			1	3	2	1		
	8						2	1	
	9								1

Figuur 2.2

Toelichting

Geef de opdracht mondeling. Kopieer vooraf het **Werkblad** en deel dit uit.

Hierbij gaat het om het vergelijken van twee datasets op basis van boxplots. Het begrip ‘kruistabel’ is wellicht nog onbekend. Ga dus na of leerlingen uit deze figuur gegevens kunnen aflezen door bijvoorbeeld vragen te stellen: “Hoeveel leerlingen hadden een 5 voor wiskunde?” en “Hoeveel leerlingen hadden een 5 voor wiskunde en ook een 5 voor science?”.

Loop na afloop de verschillende resultaatvergelijkingen na. Noem weer de begrippen ‘spreidingsbreedte’ en ‘(inter)kwartielafstand’.

Uitwerking

Wiskunde:

- het minimum is 4, het maximum is 10;
- de mediaan is 6;
- het eerste kwartiel is $Q_1 = 5$, het derde kwartiel is $Q_3 = 8$.

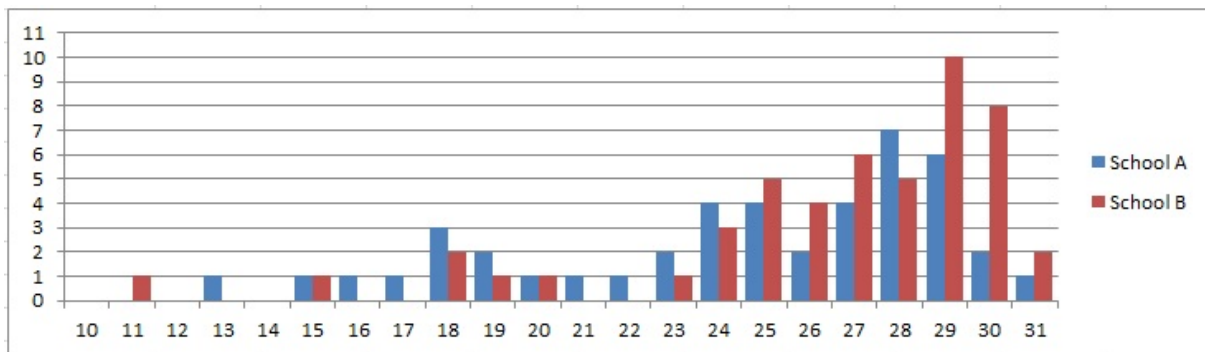
Science:

- het minimum is 4, het maximum is 9;
- de mediaan is 6;
- het eerste kwartiel is $Q_1 = 5$, het derde kwartiel is $Q_3 = 7$.

Hierbij horen twee boxplots. Voor wat betreft de laagste cijfers tot en met de 6 hebben beide vakken dezelfde spreiding. Maar voor de hoogste cijfers (dus 7 of hoger) is de spreiding bij wiskunde groter.

Opdracht 2.3

Op twee scholen voor voortgezet onderwijs zijn de aantallen leerlingen per klas geteld. Je ziet in dit staafdiagram het resultaat voor school A en school B.



Figuur 2.3

Gebruik boxplots om te bepalen welke school in het algemeen grotere klassen heeft. Vergelijk ook de beide spreidingsbreedtes en kwartielafstanden.

Toelichting

Geef de opdracht mondeling. Kopieer het diagram vooraf vanaf het **Werkblad** en deel dit uit.

Deze opdracht gaat om het trekken van een conclusie uit twee boxplots. Het gegeven diagram hebben ze eerder gezien.

— **Uitwerking** —

Gebruik eventueel een frequentietabel om de spreidingsmaten te bepalen.

School A:

Kleinste waarneming 13, $Q_1 = 21,5$, mediaan 25,5, $Q_3 = 28$ en grootste waarneming 31.

Spreidingsbreedte $31 - 13 = 18$, kwartielafstand $28 - 21,5 = 6,5$.

School B:

Kleinste waarneming 11, $Q_1 = 25$, mediaan 27,5, $Q_3 = 29$ en grootste waarneming 31.

Spreidingsbreedte $31 - 11 = 20$, kwartielafstand $29 - 25 = 4$.

Teken de bijpassende boxplots.

Bijna 75% van de klassen van school B is groter of gelijk aan de mediaan van de klassengrootten van school A.

Opdracht 2.4

Bekijk wat iedereen heeft gemaakt en heeft bedacht over het bepalen van 'spreidingsmaten' bij een verzameling data (gegevens). Verder wordt de spreiding van de data weergegeven in boxplots.

Maak een eigen overzicht van wat je hebt geleerd.

— **Toelichting** —

Loop samen met de leerlingen alle bedenksels na. Bevraag leerlingen of ze elkaars gedachtenspinnen kunnen toelichten. Samen zouden jullie naar een overzicht van de theorie moeten komen. Ieder schrijft het voor zichzelf op.

— **Uitwerking** —

Het theorieblok geeft het gewenste overzicht.

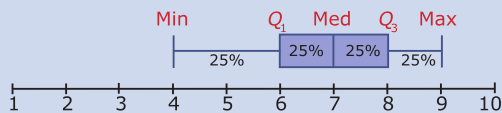
Theorie

Om te onthouden

Je kunt van een hoeveelheid gegevens een samenvatting maken, bestaande uit vijf getallen:

- de kleinste waarneming;
- het **eerste kwartiel** Q_1 , dit is de mediaan van de eerste helft waarnemingen;
- het **tweede kwartiel** is de mediaan;
- het **derde kwartiel** Q_3 , dit is de mediaan van de tweede helft waarnemingen;
- de grootste waarneming.

Deze vijf getallen zet je in een tekening langs een as en je maakt een **boxplot**. Een boxplot is een 5-getallen samenvatting van de frequentietabel.



Figuur 2.4

In een boxplot vind je twee **spreidingsmaten**:

- de **spreidingsbreedte** is de grootste waarneming min de kleinste waarneming;
- de **interkwartielafstand** is het derde kwartiel min het eerste kwartiel.

Verwerken

★ Opgave 2.1

Op 1 januari 2021 werden de volgende geboortegewichten van een groep jongens gemeten. De resultaten in grammen:

1850	2100	2500	2600	2700	2800	2900	3150	3250	3300
3300	3400	3400	3500	4000	4000	4000	4050	4090	4100
4150	4160	4180	4200	4200	4400	4500	4700	4800	4850

Tabel 2.2

Hetzelfde gebeurde bij een even grote groep meisjes:

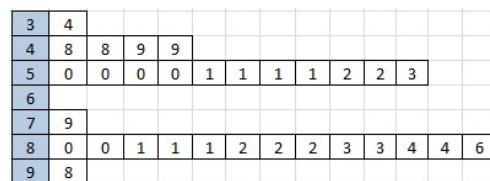
2400	2500	2500	2600	2700	2800	2800	2800	2900	2900
3000	3000	3100	3100	3200	3200	3200	3300	3300	3350
3400	3400	3500	3500	3600	3700	3800	3900	4000	4000

Tabel 2.3

- Bepaal de spreidingsbreedte bij de jongens en bij de meisjes.
- Maak de boxplots die bij deze gegevens horen.
- Hoeveel procent van de meisjes woog tussen de 2800 en 3500 gram?
- Hoeveel procent van de jongens woog meer dan het zwaarste meisje?

★ Opgave 2.2

Dit is een steelbladdiagram van de cijfers voor een scheidetoets in een vijfde klas.



Figuur 2.5

- Maak er een boxplot bij.
- Welk diagram geeft het beste beeld van de verdeling van de cijfers, het steelbladdiagram of de boxplot?

★ Opgave 2.3

Van twee honkbalteams zijn alle leeftijden genoteerd:

team 1	18	20	21	22	23	23	24	28	28
team 2	21	22	22	22	23	23	24	24	26

Tabel 2.4

- Bepaal van beide teams de gemiddelde leeftijd en de mediaan.
- Teken bijbehorende boxplots.
- Krijg je hierdoor een goede indruk van het verschil in leeftijdsopbouw tussen beide teams? Leg dat uit.

★ **Opgave 2.4**

Een van de leerlingen van de vijfde klas heeft de gewichten van al zijn klasgenoten opgeschreven. Het resultaat in kilogrammen is als volgt:

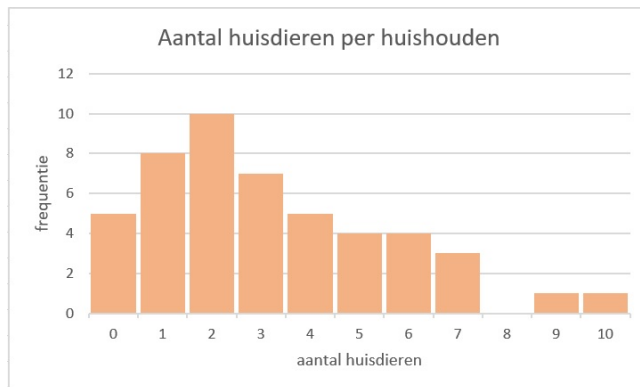
72	50	63	59	63	51	56	64	58	74
52	69	42	51	54	58	51	60	60	59
50	67	61	45	66	49	64	63	57	56

Tabel 2.5

- a Bereken bij deze gegevens de mediaan en de interkwartielafstand.
- b Hoeveel procent van deze leerlingen weegt minder dan 63 kg volgens je antwoord bij a? Is dat in werkelijkheid ook zo?

★ **Opgave 2.5**

Aan een groep kinderen uit klas 2 is gevraagd hoeveel huisdieren er in hun huishouden zijn. De resultaten zie je in het staafdiagram.

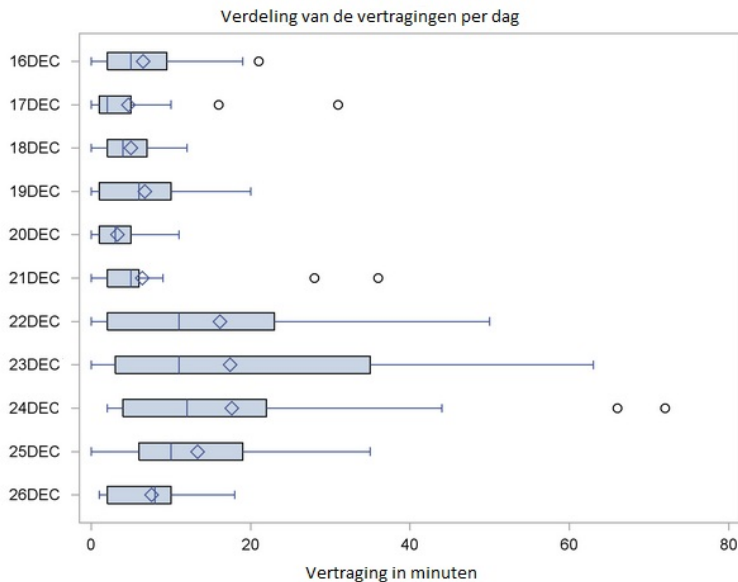


Figuur 2.6

- a Hoeveel kinderen zijn er ondervraagd?
- b Bereken het gemiddelde aantal huisdieren per huishouden.
- c Bereken de mediaan en de kwartielen.
- d Bepaal de spreidingsbreedte en de interkwartielafstand.
- e Teken de bijbehorende boxplot.
- f Hoeveel huisdieren hebben de 25% huishoudens met de meeste huisdieren?

Opgave 2.6

Op een bepaald vliegveld is een aantal opeenvolgende dagen bijgehouden hoeveel vertraging de vluchten van die dag hadden. Elke dag vinden er dezelfde vluchten plaats, in totaal 320 vluchten per dag. De echte uitschieters zijn buiten de boxplots gehouden en aangegeven door rondje. Ook is het gemiddelde aantal minuten vertraging elke dag aangegeven met een symbool in de vorm van een ruit.



Figuur 2.7

- Hoeveel uitschieters waren er op 17 december? Veranderen de mediaan en de kwartielen als deze uitschieters wel worden meegerekend?
- Waarom zie je dat bij het berekenen van het gemiddelde de uitschieters wel zijn meegeteld?
- Welke dag kende de minste vertragingen?
- Welke dag was de kleinste vertraging het grootst?
- Op welke dagen was meer dan een kwart van de vluchten meer dan 20 minuten vertraagd?
- Hoe kan het dat sommige dagen de gemiddelde vertraging nogal veel groter is dan de mediaan?

Toepassen

**

Opgave 2.7: Boxplots met een spreadsheet

Een spreadsheet kan je helpen bij het berekenen van centrum- en spreidingsmaten en het maken van boxplots. Bekijk het [Practicum](#) en het download het bestand 'Gegevens 154 Leerlingen'.

Er wordt van uit gegaan dat je wel eens eerder met Excel hebt gewerkt en al weet hoe je met cellen werkt, met formules in cellen werkt, absoluut en relatief kopiëren begrijpt, etc.

- Maak zelf een tabel met de kwartielen, de mediaan en maximum en minimum van de gehele eindcijfers voor wiskunde van deze groep leerlingen in 4 vwo.
- Maak een bijbehorend boxplot.
- Maak vervolgens afzonderlijke groepen van de jongens en de meisjes door te sorteren. Bereken de centrum- en spreidingsmaten voor deze afzonderlijke groepen en vergelijk ze met elkaar. Maak ook boxplots van de afzonderlijke groepen.

★ ★ **Opgave 2.8: Lengtes en gewichten vergelijken**

Werk met hetzelfde bestand als in de voorgaande opgave. Zet de gegevens over de lengtes van de jongens en de meisjes op een afzonderlijk werkblad. Doe hetzelfde met de gewichten van de jongens en de meisjes.

- a Maak voor de lengtes van de jongens en de meisjes afzonderlijke boxplots. Kun je conclusies trekken?
- b Maak voor de gewichten van de jongens en de meisjes afzonderlijke boxplots. Kun je conclusies trekken?

Practicum: Statistiek met een spreadsheet

Een **spreadsheet of rekenbladprogramma** kan centrum- en spreidingsmaten voor je berekenen vanuit de data.

In de volgende practica leer je werken met statistiek in een spreadsheet. Je kunt er ook boxplots in maken. Download wel eerst het bijbehorende bestand 'Gegevens 154 leerlingen'.

- [Data presenteren met behulp van MS-Excel](#)
- [Data presenteren met behulp van Open Office Calc](#)
- [Data presenteren met behulp van Google spreadsheets](#)

1.3 Klassenindeling

Inleiding

Soms heb je echt veel verschillende gegevens. Bijvoorbeeld de lengtes van 90 meisjes. Dat noem je de ruwe data. Het maken van een frequentietabel levert dan weinig overzicht op. Daarvoor kun je beter de lengtes groeperen in zogenaamde 'klassen'. Daarover gaat dit onderdeel.

Je leert in dit onderwerp

- een frequentietabel met klassenindeling en geschikte klassenbreedte maken bij een verzameling ruwe data;
- bij zo'n klassenindeling diagrammen maken;
- bij een klassenindeling het gemiddelde schatten.

Voorkennis

- (gewogen) gemiddelde, modus en mediaan van een hoeveelheid gegevens berekenen;
- werken met (relatieve) frequenties en (relatieve) frequentietabellen;
- werken met kruistabellen.

Voor de docent

Bij het onderdeel 'Klassenindeling' gaat het er om bij een serie gegevens, een dataset, een klassenindeling te maken om daarbij een overzichtelijke frequentietabel te kunnen maken. Begrippen als 'klassenmidden', 'klassenbreedte' en 'klassengrenzen' komen voorbij, net als het maken van diagrammen en het schatten van het gemiddelde bij een klassenindeling.

Het is wellicht mooier om bij de eerste twee opdrachten een set eindcijfers van enkele klassen (van een jaar geleden, of nog ouder, geanonimiseerd) van de eigen school te gebruiken.

Bij dit onderdeel hoort een Practicum waarin de leerlingen leren werken met statistiek in een spreadsheet. Laat ze dit doorwerken als er tijd voor is, het kan ook worden gebruikt voor een afsluitende opdracht van dit onderwerp.

Gewenste materialen:

- Maak vooraf een kopie van het werkblad bij de eerste, de tweede en de derde opdracht.
- Schrijfmateriaal voor op de verticale werkvlakken en eventueel plakband om de informatiebladen eraan op te hangen.

Opdracht 3.1

Dit zijn dertig rapportcijfers van klas 2A op één decimaal nauwkeurig.

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

Tabel 3.1

Om meer overzicht te krijgen ga je deze cijfers ordenen in 'klassen', groepjes cijfers die bij elkaar liggen.

Je begint met klasse $3,5- < 4,5$. Hierin komen alle cijfers vanaf 3,5 tot aan 4,5 (4,5 zelf dus niet). 3,5

en 4,5 zijn de klassengrenzen, het klassenmidden is 4.
De volgende klasse is 5,5– < 6,5, enzovoorts. Alle klassen maak je even breed.
Maak zo'n klassenindeling en de bijbehorende frequentietabel.

— **Toelichting** —

Geef de opdracht mondeling. Kopieer de gegevens vooraf vanaf het **Werkblad** en deel dit uit.
Bespreek na afloop het begrip 'klasse' en waarom voor deze klassenindeling is gekozen. (Alle cijfers in een klasse worden afgerond op hetzelfde gehele rapportcijfer.) De volgende opdracht is een direct vervolg op deze, daarin worden twee klassen vergeleken.

— **Uitwerking** —

De frequentietabel met klassenindeling wordt:

klasse	klassenmidden	frequentie
3,5– < 4,5	4	2
4,5– < 5,5	5	2
5,5– < 6,5	6	7
6,5– < 7,5	7	11
7,5– < 8,5	8	5
8,5– < 9,5	9	3
totaal		30

Tabel 3.2

Opdracht 3.2

Je krijgt nu ook de cijfers voor hetzelfde vak in klas 2B.

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

Tabel 3.3

Je wilt de resultaten van die klas vergelijken met die van klas 2A, dus je gaat daarvan ook een klassenindeling maken. Omdat er minder leerlingen in deze klas zitten, kun je de frequenties niet zo maar vergelijken met die van klas 2A. Je gaat die 'absolute frequenties' dus omrekenen naar 'relatieve frequenties', bijvoorbeeld in procenten.

Maak van deze cijfers ook een klassenindeling en reken beide kolommen met frequenties om naar kolommen met relatieve frequenties.

Hoeveel procent van de leerlingen in 2A heeft een onvoldoende (dat is een cijfer lager dan 5,5)? En hoe zit dat in 2B? Kun je een conclusie trekken?

— **Toelichting** —

Geef de opdracht mondeling. Kopieer vooraf het **Werkblad** en deel dit uit.

Hierbij gaat het om het vergelijken van twee datasets op basis van klassenindelingen.

Bespreek na afloop de voordelen en de nadelen van een klassenindeling. (Een nadeel is dat de centrum- en de spreidingsmaten niet meer exact kunnen worden berekend, maar dat er met klassenbreedte en klassenmiddens moet worden gewerkt.)

Uitwerking

klasse	klas 2A		klas 2B	
	abs. freq.	%	abs. freq.	%
3,5 < 4,5	2	6,7	1	4,0
4,5 < 5,5	2	6,7	2	8,0
5,5 < 6,5	7	23,3	12	48,0
6,5 < 7,5	11	36,7	6	24,0
7,5 < 8,5	5	16,7	4	16,0
8,5 < 9,5	3	10,0	0	0,0
totaal	30	100	25	100

Tabel 3.4

In 2A is het percentage onvoldoendes: $\frac{4}{30} \cdot 100 \approx 13,3\%$.

In 2B is dat: $\frac{3}{25} \cdot 100 = 12\%$.

In 2B zijn minder onvoldoendes gevallen, meer kun je hieruit niet concluderen.

Opdracht 3.3

Deze tabel laat de verdeling zien van de lengtes van 90 meisjes in een vierde klas. De gebruikte klassenindeling heeft een klassenbreedte van 5.

lengteklasse	klassenmidden	freq.
150– < 155	152,5	1
155– < 160	157,5	5
160– < 165	162,5	10
165– < 170	167,5	16
170– < 175	172,5	22
175– < 180	177,5	15
180– < 185	182,5	11
185– < 190	187,5	6
190– < 195	192,5	3
195– < 200	197,5	1
	totaal	90

Tabel 3.5

De werkelijke lengtes van de 90 meisjes kun je niet uit deze frequentietabel aflezen. De ruwe data zie je niet meer. Je ziet alleen de geordende data.

Als je aanneemt dat in de klasse 150– < 155 alle lengtes zitten vanaf 150,00... tot en met 154,99...

dan is het klassenmidden $\frac{150+155}{2} = 152,5$.

Maak een lijndiagram met behulp van deze klassenmiddens.

Je kunt vanuit die klassenmiddens ook het gemiddelde schatten. Laat zien, hoe.

Toelichting

Geef de opdracht mondeling. Kopieer de tabel vooraf vanaf het [Werkblad](#).

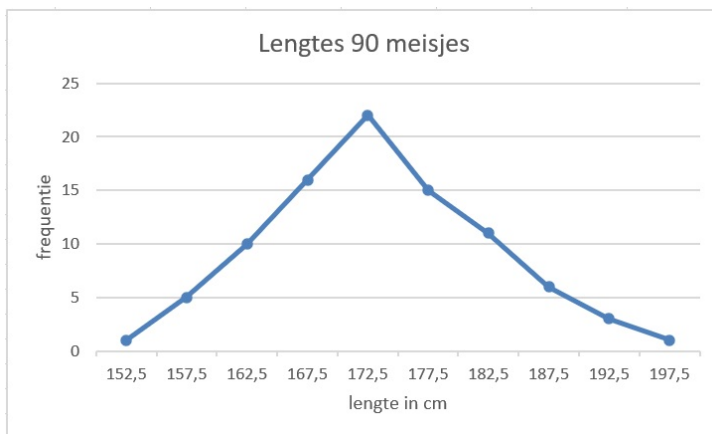
Deze opdracht gaat om het schatten van het gemiddelde bij een gegeven klassenindeling en om het tekenen van diagrammen erbij. Eventuele hulpvragen: “Wat is het belang van de klassenmiddens bij een lijndiagram?” en “En wat is het belang van de klassenmiddens bij het berekenen van het gemiddelde?”.

Bespreek na afloop ook hoe je hierbij een staafdiagram zou kunnen maken en welke rol de klassenmiddens en de klassengrenzen dan hebben.

Uitwerking

Hier zie je het lijndiagram bij de lengte van de meisjes.

De dikke punten zijn de frequenties die boven het klassenmiddens liggen. Deze dikke punten verbind je.



Figuur 3.1

Je schat het gemiddelde van alle 90 meisjes door elk klassenmiddens te vermenigvuldigen met de bijbehorende frequentie, al deze uitkomsten bij elkaar op te tellen, en vervolgens te delen door het totaal aantal meisjes. Je komt dan uit op 173,4 cm.

Opdracht 3.4

Bekijk wat iedereen heeft gemaakt en heeft bedacht over het maken van een ‘klassenindeling’ bij een verzameling data (gegevens). Bekijk ook goed wat de klassengrenzen en de klassenmiddens zijn en hoe je daarmee vanuit een klassenindeling het gemiddelde kunt schatten.

Maak een eigen overzicht van wat je hebt geleerd.

Toelichting

Loop samen met de leerlingen alle bedenksels na. Bevraag leerlingen of ze elkaars gedachtenspinsels kunnen toelichten. Samen zouden jullie naar een overzicht van de theorie moeten komen. Ieder schrijft het voor zichzelf op.

Uitwerking

Het theorieblok geeft het gewenste overzicht.



Theorie

Om te onthouden

Een groep gegevens (lieft in de vorm van getallen) noem je **ruwe data**, ruwe gegevens. Erg overzichtelijk zijn de ruwe data meestal niet.

Dan ga je ze ordenen in **klassen**, dat zijn groepjes getallen die dicht bij elkaar liggen.

Je begint met het kiezen van de **klassengrenzen**. Normaal gesproken kies je alle klassen even breed.

Het verschil van de klassengrenzen van een klasse heet de **klassenbreedte**.

Het **klassenmidden** is meestal het gemiddelde van de twee klassengrenzen.

Het aantal getallen dat in een bepaalde klasse komt is de **absolute frequentie** van de klasse. Vaak zijn **relatieve frequenties** handiger. De relatieve frequentie bepaal je door de frequentie te delen door het totaal. Ook kun je het **percentage** berekenen, je vermenigvuldigt dan de relatieve frequentie met 100.

Nu heb je **geordende data**, de cijfers van de klas zijn overzichtelijk weergegeven in een **klassenindeling**. Het gemiddelde kun je nu alleen nog maar schatten met behulp van de klassenmiddens.

Verwerken

★ Opgave 3.1

Geef in de volgende situaties aan of een klassenindeling maken mogelijk/zinvol is, leg uit waarom. Zo ja, geef aan wat jij als eerste klasse zou kiezen.

- Je hebt op de huishoudbeurs aan 500 bezoekers hun leeftijd gevraagd. De jongste bezoeker was 17 en de oudste was 81.
- Je hebt aan die zelfde 500 bezoekers van de huishoudbeurs hun lievelingskleur gevraagd. Tussen de antwoorden die de bezoekers gaven zaten 12 verschillende kleuren.
- Je hebt aan die zelfde 500 bezoekers van de huishoudbeurs gevraagd hoe vaak ze nu in totaal de huishoudbeurs al bezocht hebben. Het laagste aantal was 1 keer en het hoogste aantal was 15 keer.

★ Opgave 3.2

Johan gooit zestig keer met vijf dobbelstenen en noteert telkens de som van de vijf aantallen ogen. Zijn scores zijn als volgt:

16	22	21	16	20	24	20	23	16	22
17	15	17	26	23	20	17	19	21	23
17	22	18	14	18	15	16	18	17	18
20	22	18	18	25	14	13	12	15	17
22	14	18	20	12	20	15	18	15	24
17	12	23	14	11	14	21	16	18	5

Tabel 3.6

- Wat was het grootste getal dat hij had kunnen gooien? Leg uit.
- Bereken de gemiddelde score in één decimaal nauwkeurig.
- Maak een klassenindeling van de scores. Neem als klassen $5- < 8$ en $8- < 11$, enzovoorts.
- Welke scores horen tot de klasse $5- < 8$? Wat is dus het klassenmidden? En waarom kun je nu niet gewoon het gemiddelde van de klassengrenzen nemen?
- Bereken nu met behulp van de klassenindeling de gemiddelde score.
- In hoeveel procent van de worpen scoorde Johan meer dan 13 maar minder dan 23?
- Waarom zou score 18 zoveel vaker voorkomen dan score 28?

★ Opgave 3.3

Je ziet hier een klassenindeling van de gewichten van een groep personen.

- Welk klassenmidden heeft de klasse $45- < 50$? Licht je antwoord toe.
- Bereken het gemiddelde gewicht in één decimaal nauwkeurig.
- Hoeveel procent van deze groep mensen weegt minder dan 70 kg?

gewichtsklasse	frequentie
45- < 50	1
50- < 55	5
55- < 60	8
60- < 65	15
65- < 70	19
70- < 75	9
75- < 80	3

Tabel 3.7

★ **Opgave 3.4**

Bij een cabaretvorstelling is aan 44 bezoekers hun leeftijd gevraagd. Het resultaat is als volgt:

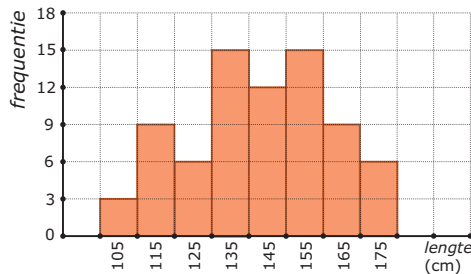
31	29	18	16	24	47	12	32	52	10	26
22	12	53	49	25	21	59	51	32	16	27
18	29	31	37	17	42	31	39	19	33	28
44	16	29	22	17	19	54	27	30	28	51

Tabel 3.8

- a Maak een klassenindeling met als klassen $10- < 15$, enzovoorts. Voeg de absolute frequenties toe.
- b Welke leeftijden horen tot de klasse $50- < 55$? Wat is dus het klassenmidden?
- c Schat met behulp van de klassenindeling de gemiddelde leeftijd van deze bezoekers.
- d Hoeveel procent van de ondervraagden was jonger dan 30?

★ **Opgave 3.5**

Je ziet hier een verdeling van de lengtes van de vrouwelijke leerlingen van een balletschool. Op de horizontale as staan de klassenmiddens in cm.



Figuur 3.2

- a Schrijf de eerste klasse van de indeling op.
- b Van hoeveel leerlingen is de lengte gemeten?
- c Schat met behulp van de klassenindeling de gemiddelde lengte van deze leerlingen in cm nauwkeurig.
- d Maak een nieuwe klassenindeling met klassen $100- < 120$, enzovoorts. Bereken opnieuw de gemiddelde lengte van deze leerlingen, maar nu vanuit de nieuwe klassenindeling. Verklaar het verschil.

★ **Opgave 3.6**

Op 1 januari 2021 werden de volgende geboortegewichten van een groep jongens gemeten. De resultaten in grammen:

1850	2100	2500	2600	2700	2800	2900	3150	3250	3300
3300	3400	3400	3500	4000	4000	4000	4050	4090	4100
4150	4160	4180	4200	4200	4400	4500	4700	4800	4850

Tabel 3.9

Hetzelfde gebeurde bij een even grote groep meisjes:

2400	2500	2500	2600	2700	2800	2800	2800	2900	2900
3000	3000	3100	3100	3200	3200	3200	3300	3300	3350
3400	3400	3500	3500	3600	3700	3800	3900	4000	4000

Tabel 3.10

Omdat je de geboortegewichten van deze twee groepen wilt vergelijken maak je een klassenindeling en de bijbehorende frequentietabellen. Neem klassen met een breedte van 500 gram. De eerste klasse begint bij 1500.

- a Maak een geschikte klassenindeling met bijbehorende absolute frequenties.
- b Schat de gemiddelden van beide groepen met behulp van de klassenindeling bij a.
- c Hoeveel procent van de jongens woog meer dan het bij b gevonden gemiddelde van de meisjes?
- d Hoeveel procent van de jongens woog meer dan het zwaarste meisje?

Toepassen

★ ★ Opgave 3.7: Lengtes in klassen in een spreadsheet

Een spreadsheet (rekenbladprogramma) kan centrummaten en spreidingsmaten voor je berekenen vanuit de ruwe data, je hebt daarvoor geen frequentietabellen nodig. Maar frequentietabellen zijn handig bij het maken van diagrammen. En vooral als je over veel gegevens beschikt is groeperen in klassen erg handig. Bekijk het **Practicum** en het bestand "Gegevens 154 Leerlingen".

Denk er wel om dat je op een pc dit bestand eerst moet downloaden. Er wordt van uit gegaan dat je wel eens eerder met een spreadsheet hebt gewerkt en al weet hoe je met cellen werkt, met formules in cellen werkt, absoluut en relatief kopiëren begrijpt, etc.

- a Maak zelf de frequentieverdeling van de lengtes van deze groep leerlingen in 4 vwo.
- b Maak een staafdiagram.
- c Maak vervolgens afzonderlijke groepen van de jongens en de meisjes door te sorteren. Maak frequentietabellen voor deze afzonderlijke groepen en vergelijk ze met elkaar.

★ ★ Opgave 3.8: Gemiddelden schatten

Werk met hetzelfde bestand als in de voorgaande opgave. Je hebt de gegevens over de lengtes van de jongens en de meisjes op een afzonderlijk werkblad gezet. Doe hetzelfde met de gewichten van de jongens en de meisjes.

- a Maak voor de lengtes van de jongens en de meisjes afzonderlijke frequentietabellen en schat met behulp daarvan hun gemiddelde lengtes. Kun je conclusies trekken?
- b Maak voor de gewichten van de jongens en de meisjes afzonderlijke frequentietabellen en schat met behulp daarvan hun gemiddelde lengtes. Kun je conclusies trekken?

Practicum: Statistiek met een spreadsheet

Een **spreadsheet of rekenbladprogramma** kan centrummaten en spreidingsmaten voor je berekenen vanuit de ruwe data, je hebt daarvoor geen frequentietabellen nodig. Maar frequentietabellen zijn handig bij het maken van diagrammen. En vooral als je over veel gegevens beschikt is groeperen in klassen erg handig.

In de volgende practica leer je werken met statistiek in een spreadsheet:

- [Data presenteren met behulp van MS-Excel](#)
- [Data presenteren met behulp van Open Office Calc](#)
- [Data presenteren met behulp van Google spreadsheets](#)

1.4 Schattingen

Inleiding

Je hebt leren werken met klassenindelingen en de bijbehorende frequentietabellen. Voordeel ervan is het krijgen van enig overzicht als je met veel ruwe data te maken hebt. Nadeel is dat je alle centrummaten en spreidingsmaten alleen nog maar kunt schatten.

Je leert in dit onderwerp

- bij een klassenindeling de centrummaten en de spreidingsmaten schatten;
- bij een klassenindeling een boxplot maken.

Voorkennis

- (gewogen) gemiddelde, modus en mediaan van een hoeveelheid gegevens berekenen;
- een frequentietabel met klassenindeling en geschikte klassenbreedte maken bij een verzameling ruwe data;
- werken met kruistabellen.

Voor de docent

Bij het onderdeel 'Schatten' gaat het er om bij een serie gegevens, een dataset, gegeven door een klassenindeling het gemiddelde, de mediaan, de kwartielen en de grenswaarden van de boxplot te schatten.

Het is wellicht mooier om bij de eerste opdracht een set eindcijfers van enkele klassen (van een jaar geleden, of nog ouder, geanonimiseerd) van de eigen school te gebruiken.

Gewenste materialen:

- Maak vooraf een kopie van het werkblad bij de eerste en de tweede opdracht.
- Schrijfmateriaal voor op de verticale werkvlakken en eventueel plakband om de informatiebladen eraan op te hangen.

Opdracht 4.1

Dit is een klassenindeling van de rapportcijfers van klas 2A.

klasse	klassenmidden	frequentie
3,5– < 4,5	4	2
4,5– < 5,5	5	2
5,5– < 6,5	6	7
6,5– < 7,5	7	11
7,5– < 8,5	8	5
8,5– < 9,5	9	3
totaal		30

Tabel 4.1

De cijfers die op één decimaal nauwkeurig zijn gehaald kun je niet meer uit deze tabel aflezen. Dus je kunt ook niet langer het gemiddelde op één decimaal nauwkeurig berekenen, je kunt het alleen nog schatten. En de mediaan in één decimaal nauwkeurig bepalen is ook lastig.

Bereken bij deze klassenindeling een schatting van het gemiddelde in één decimaal nauwkeurig. Teken ook een bijpassend boxplot met de grenzen, de kwartielen en de mediaan in één decimaal nauwkeurig.

Toelichting

Geef de opdracht mondeling. Kopieer de gegevens vooraf vanaf het **Werkblad** en deel dit uit.

Mogelijke hulpvragen: “Welke getallen ga je gebruiken om het gemiddelde te schatten?”, “Welke grenzen heeft de boxplot op één decimaal nauwkeurig?”, “In welke klasse zit de mediaan?” en “Hoe ga je die mediaan op één decimaal nauwkeurig bepalen? Welke aanname moet je dan doen?” (en idem voor de kwartielen).

Uitwerking

Het gemiddelde schat je nu vanuit de klassenmiddens: $(4 \cdot 2 + 5 \cdot 2 + 6 \cdot 7 + 7 \cdot 11 + 8 \cdot 5 + 9 \cdot 3)/30 = 6,8$.

Voor een boxplot heb je vijf getallen nodig: de laagste waarde, de hoogste waarde, het eerste kwartiel, de mediaan en het derde kwartiel. Nu kun je deze waarden gemakkelijk bepalen als je gewoon de klassenmiddens (afgeronde, gehele cijfers) gebruikt. Maar dit boxplot past niet goed bij de werkelijke cijfers.

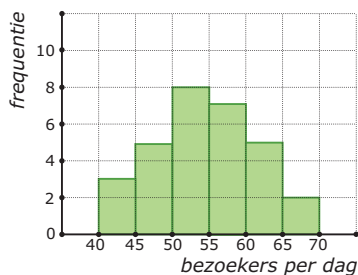
Je kunt de vijf gezochte waarden echter schatten:

- De laagste waarde schat je 3,5 en de hoogste is 9,5.
- Voor de mediaan heb je het 15e en 16e getal nodig. Die zitten beide in de klasse $6,5- < 7,5$. Daar zitten 10 getallen in. Neem aan dat die gelijkmatig oplopen vanaf 6,5. Dan verschillen ze 0,1 van elkaar. Dan is het 15e getal $6,5 + 3 \cdot 0,1 = 6,8$ en het 16e getal $6,5 + 4 \cdot 0,1 = 6,9$. De mediaan is dus 6,85.
- Het eerste kwartiel is het 8e getal. Dat schat je net zo als de mediaan. Je vindt ongeveer 5,9.
- Het derde kwartiel is het 23e getal. En dat schat je als 7,5.

Nu is de gewenste boxplot te tekenen.

Opdracht 4.2

Dit staafdiagram geeft voor de maand juni de aantallen bezoekers van een klein museum.



Figuur 4.1

Hoe berekenen je hierbij het gemiddelde aantal bezoekers per dag?

Toelichting

Geef de opdracht mondeling. Kopieer vooraf het **Werkblad** en deel dit uit.

Hulpvragen: “Welke getallen zijn nu de klassenmiddens?”.

Uitwerking

Je kunt het gemiddelde schatten met behulp van de klassenmiddens. Daarbij moet je rekening houden met het feit dat het hier om gehele aantallen gaat.

De klasse $40- < 45$ bestaat uit de aantallen 40, 41, 42, 43 en 44. Het midden daarvan is 42.

Op dezelfde wijze vind je de andere klassenmiddens 47, 52, enzovoorts. Nu kun je het gemiddelde schatten.

Je krijgt ongeveer: $(42 \cdot 3 + 47 \cdot 5 + 52 \cdot 8 + 57 \cdot 7 + 62 \cdot 5 + 67 \cdot 2)/(3 + 5 + 8 + 7 + 5 + 2) = \frac{1620}{30} = 54$ bezoekers per dag.

Bij het berekenen van een klassenmidden moet je bedenken of echt alle getallen tussen de ondergrens en de bovengrens voor kunnen komen, of alleen maar de gehele getallen!



Opdracht 4.3

Bekijk wat iedereen heeft gemaakt en heeft bedacht over het maken van een schatting van een gemiddelde en het maken van een boxplot bij een verzameling data (gegevens) die alleen als frequentieverdeling is gegeven.

Maak een eigen overzicht van wat je hebt geleerd.

Toelichting

Loop samen met de leerlingen alle bedenksels na. Bevraag leerlingen of ze elkaars gedachtenspingsels kunnen toelichten. Samen zouden jullie naar een overzicht van de theorie moeten komen. Ieder schrijft het voor zichzelf op.

Uitwerking

Het theorieblok geeft het gewenste overzicht.



Theorie

Om te onthouden

Vanuit een klassenindeling kun je van de verdeling van de cijfers op één decimaal de centrummaten (modus, mediaan, gemiddelde) en de spreidingsmaten (spreidingsbreedte, kwartielafstand) niet meer precies berekenen.

In plaats daarvan gebruik je:

- de **modale klasse**, dat is de klasse met de hoogste frequentie;
- de klassenmiddens om het **gemiddelde te schatten**;
- het aantal klassen maal (keer) de klassenbreedte voor de **spreidingsbreedte** of **variatiebreedte**.

Om een **boxplot** te maken moet je de hoogste en de laagste waarde, kwartielen en de mediaan schatten.

Verwerken

★ Opgave 4.1

In een klas zijn voor een toets de volgende cijfers behaald:

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

Tabel 4.2

- Bepaal de variatiebreedte.
- Als je een staafdiagram van deze gegevens wilt maken is het verstandig om eerst een klassenindeling te maken. Waarom is dat?
- Maak een klassenindeling van de resultaten en teken het staafdiagram. Welke klassen liggen hier voor de hand?
- Bepaal de modale klasse.
- In welke klasse ligt de mediaan? Hoeveel wordt de mediaan naar schatting?
- Hoeveel wordt het eerste kwartiel naar schatting?

★ Opgave 4.2

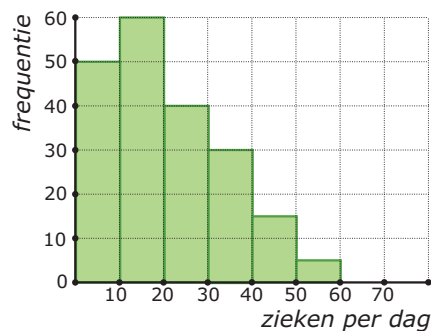
Bedenk in de volgende situaties wat het klassenmidden is van de gegeven klasse.

- De cijfers die klas 2C voor scheikunde heeft gehaald. Klasse 7– < 8.
- Het aantal zieke leerlingen per schooldag. Klasse 10– < 15.
- De lichaamslengtes van een groep kinderen. Klasse 160– < 170.
- Het aantal bezoekers per dag van een winkeltje. Klasse 30– < 45.

★ Opgave 4.3

Op een school is gedurende een jaar elke schooldag bijgehouden hoeveel leerlingen er ziek waren. Je ziet de gegevens hier verwerkt in een staafdiagram. De laatste klasse bevat in feite alle dagen dat het aantal zieken 50 of meer was.

- Hoeveel schooldagen waren er dat jaar?
- Schat de mediaan.
- Schat het gemiddelde aantal zieke leerlingen per dag in gehelen. Neem voor de laatste klasse als klassenmidden 55.
- Je hebt voor de laatste klasse als klassenmidden 55 genomen. Voor welke centrummaat is die keuze van belang? Licht je antwoord toe.



Figuur 4.2

★ **Opgave 4.4**

De lengtes in cm van 60 leerlingen zijn in klassen ingedeeld. Je kunt nu het gemiddelde alleen nog schatten want je weet de ruwe data niet.

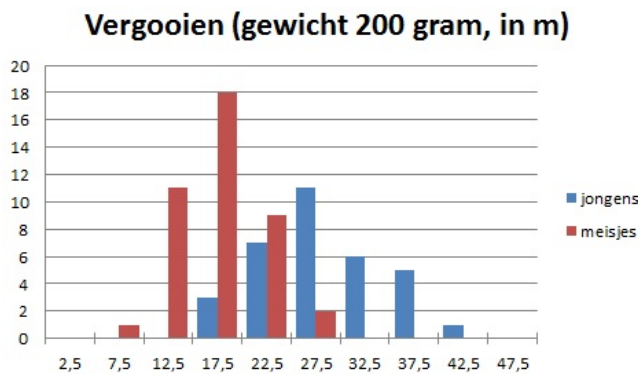
- a Bepaal de modale klasse.
- b Hoe schat je het gemiddelde? Bereken deze schatting.
- c Schat de mediaan.
- d Schat de kwartielen en teken een bijpassend boxplot.

lengteklasse	frequentie
150– < 160	3
160– < 170	17
170– < 180	34
180– < 190	4
190– < 200	2
totaal	60

Tabel 4.3

★ **Opgave 4.5**

Hier zie je in één staafdiagram de resultaten van het gooien met een 200 grams kogel van een groep van 74 brugklasleerlingen. Er zijn afzonderlijke diagrammen voor de jongens en de meisjes. Alle geworpen afstanden zijn tot op 0,5 m nauwkeurig gemeten. Elke klasse is door het klassenmidden aangegeven.



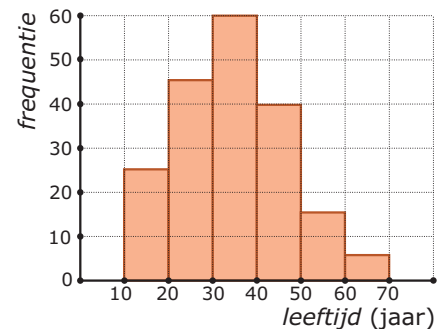
Figuur 4.3

- a Maak bij beide staafdiagrammen een boxplot.
- b Maak een schatting van het gemiddelde van de jongens en de meisjes afzonderlijk.
- c Kun je concluderen dat jongens van die leeftijd verder gooien dan meisjes? Licht je antwoord toe.

★ **Opgave 4.6**

In het staafdiagram zie je de leeftijdsverdeling van een groep reizigers die met bussen op weg is naar de Costa del Sol.

- a Maak bij het staafdiagram een frequentieverdeling. Neem dezelfde klassen als in het staafdiagram.
- b Bepaal de spreidingsbreedte.
- c Schat de gemiddelde leeftijd van de reizigers in gehele jaren.
- d Geef een schatting van de mediaan. Beredeneer je schatting.
- e Geef een schatting van de kwartielen en bereken de kwartielafstand.
- f Hoeveel procent van de buspassagiers was jonger dan de gemiddelde leeftijd?
- g De reisorganisatie geeft korting aan alle reizigers die ouder zijn dan 55 jaar of jonger dan 15 jaar. Geef een schatting van het percentage reizigers dat korting kreeg.
- h Welke klasse is de modale klasse?
- i Tijdens de vakantie zijn twaalf van de reizigers jarig. Verandert hierdoor de gemiddelde leeftijd in gehelen?



Figuur 4.4

Toepassen

★ ★ Opgave 4.7: Uren huiswerk per week

Je hebt de belangrijkste vaardigheden met MS-Excel in de voorgaande onderdelen voorbij zien komen. Je hebt centrummaten en spreidingsmaten berekend (dat doe je het liefst vanuit de ruwe data) en je hebt diagrammen getekend. Voor een boxplot gebruik je ook het liefst de ruwe data, maar voor een staafdiagram of een lijndiagram of een cirkeldiagram werk je meestal met een klassenindeling. In de volgende opgaven werk je met het bestand [Gegevens154Leerlingen.xls](#).

In dit bestand vind je een tabel met het aantal uren huiswerk per week. Je gaat de jongens en de meisjes van deze vierde klas met elkaar vergelijken.

Denk er wel om dat je op een pc het bijbehorende Excel-bestand eerst moet downloaden.

- a Maak een geschikte frequentieverdeling van het aantal uren dat de jongens aan hun huiswerk besteden. Doe dit ook voor de meisjes.
- b Maak met Excel een passend lijndiagram van de jongens en meisjes in één figuur.
- c Probeer een conclusie te trekken.

★ ★ Opgave 4.8: Uren huiswerk per week en boxplots

Werk met hetzelfde Excel-bestand als in de voorgaande opgave.

- a Maak nu twee boxplots naast elkaar voor het aantal uren huiswerk dat de jongens per week besteden en het aantal uren huiswerk dat de meisjes per week besteden.
- b Probeer opnieuw conclusies te trekken, betrek ook de gemiddelden erbij.

1.5 Statistische uitspraken

Inleiding

Je hebt inmiddels allerlei tabellen en diagrammen leren maken om data overzichtelijker weer te geven. Die data krijg je door statistisch onderzoek. Hoog tijd om daar eens een eerste kennismaking mee te hebben.



Figuur 5.1

Je leert in dit onderwerp

- een statistische uitspraak herkennen;
- bij een statistisch onderzoek de begrippen populatie en steekproef herkennen;
- aangeven of een steekproef representatief is;
- een eenvoudig statistisch onderzoek opzetten.

Voorkennis

- (gewogen) gemiddelde, modus en mediaan van een hoeveelheid gegevens berekenen;
- een frequentietabel met klassenindeling en geschikte klassenbreedte maken bij een verzameling ruwe data;
- werken met diverse tabellen (waaronder kruistabellen) en er verschillende diagrammen bij maken;
- bij een klassenindeling de centrummaten en de spreidingsmaten schatten;
- bij een klassenindeling een lijndiagram, een staafdiagram en een boxplot maken.

Voor de docent

Bij het onderdeel 'Uitspraken doen' maken de leerlingen voor het eerst kennis met wat een statistisch onderzoek inhoudt. Ze ontdekken wanneer een uitspraak statistisch onderzoek nodig heeft, wat de 'populatie' en wat een 'representatieve steekproef' is.

Het is mooi om dit onderwerp als geheel af te sluiten met een statistisch onderzoek(je).

Gewenste materialen:

- Maak vooraf een kopie van het werkblad bij de eerste opdracht.
- Schrijfmateriaal voor op de verticale werkvlakken en eventueel plakband om de informatiebladen eraan op te hangen.

Opdracht 5.1

Niet elke bewering is een statistische uitspraak, een uitspraak die je met behulp van statistiek kunt nagaan. Geef van elk van de volgende uitspraken aan of het een statistische uitspraak betreft.

1. "40% van de leerlingen in klas 2C van school X is een meisje."
2. "40% van de Nederlanders heeft bloedgroep O-positief."
3. "85% van alle jongeren tussen 12 en 18 jaar internet dagelijks."
4. "Ik heb zojuist in 360 worpen met twee dobbelstenen precies 60 keer 7 ogen gegooid."
5. "Als je met twee dobbelstenen gooit zul je in 16,7% van alle worpen 7 ogen gooien."
6. "Jongens kunnen verder gooien dan meisjes."



Toelichting

Geef de opdracht mondeling. Kopieer de uitspraken vooraf vanaf het **Werkblad** en deel dit uit.

Mogelijke hulpvragen: “Waarom zou je een uitspraak ‘statistisch’ noemen?” en “Wat is kenmerkend aan een uitspraak die beslist niet van statistiek afhangt?”.

Uitwerking

1. Geen statistische uitspraak, dit kun je gewoon opzoeken in de schooladministratie en dan het percentage uitrekenen.
2. Dit is beslist een statistische uitspraak. Je kunt dit alleen nagaan door gegevens van een deel van de Nederlanders te verzamelen. Het is onmogelijk om in een redelijk kort tijdsbestek alle Nederlanders te vragen, bovendien is dit aantal niet constant.
3. Statistische uitspraak. Je kunt dit alleen bepalen door gegevens te verzamelen van een deel van deze jongeren. Dat deel moet dan wel een goed beeld geven van de groep jongeren als geheel.
4. Geen statistische uitspraak, dit is gewoon een meting.
5. Statistische uitspraak, dit percentage kun je wel beredeneren, maar of het in de praktijk uitkomt is maar de vraag. Het percentage van 16,7% zal heel vaak niet precies uitkomen. Je gaat dan heel vaak proberen om te kijken of het percentage worpen waar 7 uit komt in de buurt van de 16,7 komt.
6. Statistische uitspraak, je moet dit nagaan door veel gegevens te verzamelen waarbij je jongens en meisjes vergelijkt.

Opdracht 5.2

Stel je eens voor dat je zou willen weten hoeveel procent van de Nederlanders kleurenblind is.

Even zoeken op internet en je weet het toch? Nee, natuurlijk niet...

Niemand weet dat precies, want het is onmogelijk om iedere Nederlander te testen op kleurenblindheid en bovendien komen er voortdurend mensen bij en sterven er mensen. Dus dit percentage kun je nooit precies weten!

Hiervoor is een statistisch onderzoek nodig met een representatieve steekproef.

Omschrijf wat de te onderzoeken groep (de ‘populatie’) is en hoe je een steekproef, die een goede vertegenwoordiging van de populatie is (een ‘representatieve steekproef’, zou kunnen samenstellen. Beschrijf hoe je dit onderzoek zou willen opzetten.

Toelichting

Geef de opdracht mondeling.

Hulpvragen: “Waarmee moet je rekening houden als je de populatie beschrijft?”, “Kan kleurenblindheid nog afhangen van specifieke groepen Nederlanders?” en “Wanneer heb je een goede vertegenwoordiging van alle Nederlanders?”.

Vergelijk na afloop de antwoorden van de verschillende groepen en probeer tot een geschikte opzet voor een statistisch onderzoek(je) te komen.

Uitwerking

De totale populatie is eigenlijk: alle Nederlanders op een zeker tijdstip.

De beste representatieve steekproef is een willekeurige groep van zo'n 2000 Nederlanders. Alleen heb je dan wel alle groeperingen binnen de NL-bevolking te pakken? En hoe bereik je die mensen? En geven ze dan ook echt en eerlijk antwoord?

Verder een eigen antwoord dat zou moeten voorbereiden op een daadwerkelijk onderzoekje.



Opdracht 5.3

Bekijk wat iedereen heeft gemaakt en heeft bedacht over het opzetten van een statistisch onderzoek. Met name het vinden van een goede te onderzoeken statistische uitspraak en de begrippen 'populatie' en 'representatieve steekproef' zijn van belang.

Maak een eigen overzicht van wat je hebt geleerd.

Toelichting

Loop samen met de leerlingen alle bedenksels na. Bevraag leerlingen of ze elkaars gedachtenspinsels kunnen toelichten. Samen zouden jullie naar een overzicht van de theorie moeten komen. Ieder schrijft het voor zichzelf op. Het mooiste is om dit onderwerp af te sluiten met een statistisch onderzoekje waarin de voorgaande kennis (frequentietabellen, centrum- en spreidingsmaten, diagrammen en de mogelijkheid tot het trekken van conclusies) zichtbaar wordt.

Uitwerking

Het theorieblok geeft het gewenste overzicht.



Theorie

Om te onthouden

Een **statistische uitspraak** over een groep is een uitspraak die is gebaseerd op gegevens van een deel van de groep.

Soms is een groep zo groot dat je niet even ieder lid van de groep kunt bekijken/bevragen.

Een vraag over zo'n grote groep beantwoord je met behulp van **statistisch onderzoek**.

Je bekijkt/bevraagt een beperkt aantal leden van de groep en verzamelt daarvan gegevens. Je noemt dat een **steekproef** trekken uit de totale **populatie** (bijvoorbeeld alle Nederlanders). Zo'n steekproef moet een goede vertegenwoordiging van de totale populatie zijn, **representatief** zijn voor de populatie. Pas dan kun je met enige betrouwbaarheid zeggen dat de gevonden gegevens ook iets zeggen over de hele groep. Meer kun je met statistiek niet doen...



Verwerken

★ Opgave 5.1

Bij statistisch onderzoek wil je een uitspraak doen over een grote groep (de populatie) op grond van de resultaten van een steekproef uit die groep.

- Waar moet een goede steekproef aan voldoen?
- Welke twee soorten statistisch onderzoek ken je op dit moment? Geef van beide een voorbeeld.

★ Opgave 5.2

Als je onderzoek wilt doen naar de hobby's van Nederlanders van 18 jaar en ouder, dan zijn belangrijke kenmerken van de populatie: geslacht, leeftijd, regio, en opleiding.

Bedenk bij elk van de volgende uitspraken om welke populatie het gaat en welke kenmerken voor die populatie belangrijk zijn om na te gaan of de steekproef representatief is.

- “De boeren in Europa die kippeneieren produceren zitten in een crisis.”
- “Er zijn veel mensen in Nederland van plan om komend jaar hun huis te gaan verbouwen.”
- “Nederlandse jongens zijn beter in atletiek dan Nederlandse meisjes.”
- “Vrouwen gaan minder vaak naar de kapper dan mannen.”

★ Opgave 5.3

Je bent journalist en wilt een artikel schrijven over studenten. Je bent vooral nieuwsgierig naar mensen die een vwo diploma behaald hebben en daarna een hbo studie gaan doen in plaats van naar de universiteit te gaan. Je hebt al wat rondgevraagd en krijgt het vermoeden dat vooral veel meisjes deze keuze maken, en dat veel minder jongens dat doen.

Je hebt inmiddels met 25 studenten gesproken die een hbo opleiding zijn gestart nadat ze hun vwo diploma gehaald hebben. 15 van deze studenten zijn vrouw en 10 zijn man. Je schrijft als kop boven je artikel: “Meisjes met een vwo diploma kiezen veel vaker voor een hbo opleiding dan jongens met een vwo diploma.”

- Is dit een statistische uitspraak? Leg uit waarom wel of niet.
- Via de website van het [Centraal Bureau voor de Statistiek](#) kun je aan meer informatie komen voor je artikel. Bekijk de statistieken over instroom, doorstroom, en uitstroom in het onderwijs.

Je kunt kiezen bij Geslacht voor het totaal aantal, het aantal mannen of het aantal vrouwen.

Je kunt kiezen bij Periode voor meerdere jaren.

Je kunt kiezen bij Onderwijspositie basisjaar voor 6 vwo met/zonder diploma.

Hoe zou je deze gegevens gebruiken voor je artikel?

Toepassen

★★ Opgave 5.4: Een eigen onderzoeksvraag bedenken

Bekijk het [Practicum](#) en het bestand ‘Gegevens 154 Leerlingen’. Dit bestand bevat informatie over een groep 4 vwo leerlingen. Deze gegevens ga je gebruiken om een statistische onderzoeksvraag te beantwoorden die jij zelf bedenkt.

De enige voorwaarde is dat het een onderzoek moet zijn waarbij je ‘op zoek gaat naar het percentage van een bepaalde populatie dat een zekere eigenschap heeft.’

- Bekijk de gegevens in het bestand en formuleer een goede statistische onderzoeksvraag.
- Gebruik de gegevens uit het bestand om één duidelijke grafiek te maken die past bij je onderzoeksvraag.
- Trek aan de hand van je grafiek conclusies en beantwoord je onderzoeksvraag.



★ ★

Opgave 5.5: Een eigen statistisch onderzoek doen

Het is nu tijd om zelf een statistisch onderzoek te doen. Kies een onderwerp gebaseerd op de twee soorten statistische onderzoeksvragen waar je tot nu toe mee hebt kennism gemaakt. Bedenk vervolgens hoe je het onderzoek kunt uitvoeren, hoe je de gegevens kunt verzamelen en overzichtelijk weergeven. En bedenk tenslotte hoe je conclusies kunt trekken en verwoorden.

Hieronder kun je zien welke stappen je moet zetten om een goed resultaat te krijgen.

- a Bedenk een eigen statistische onderzoeksvraag.
- b Stel een plan van aanpak op. Daarin beschrijf je hoe de steekproef wordt samengesteld en hoe je het onderzoek gaat aanpakken.
- c Voer je onderzoek uit en verzamel alle gegevens (bijvoorbeeld in MS-Excel).
- d Bedenk een manier om je gegevens goed te presenteren, denk aan tabellen en grafieken.
- e Trek een conclusie.
- f Kijk terug op je onderzoek en op de aanpak. Wat had beter gekund?

Practicum: Statistiek met een spreadsheet

Een **spreadsheet of rekenbladprogramma** kan centrummaten en spreidingsmaten voor je berekenen vanuit de ruwe data, je hebt daarvoor geen frequentietabellen nodig. Maar frequentietabellen zijn handig bij het maken van diagrammen. En vooral als je over veel gegevens beschikt is groeperen in klassen erg handig.

In de volgende practica leer je werken met statistiek in een spreadsheet:

- **Data presenteren met behulp van MS-Excel**
- **Data presenteren met behulp van Open Office Calc**
- **Data presenteren met behulp van Google spreadsheets**

1.6 Totaalbeeld

Samenvatten

Begrippenlijst

- frequentietabel — centrummaat — modus — mediaan — gemiddelde
- boxplot — kwartiel — spreidingsbreedte — (inter)kwartielafstand
- klassenindeling — klassengrenzen, klassenmidden, klassenbreedte
- schatten
- statistisch onderzoek — populatie — steekproef — representatief

Activiteitenlijst

- gegevens samenvatten in frequentietabellen en beschrijven met behulp van centrummaten;
- gegevens samenvatten in frequentietabellen en beschrijven met behulp van spreidingsmaten en boxplots;
- klassenindelingen gebruiken, het gemiddelde schatten;
- centrummaten en spreidingsmaten schatten vanuit een klassenindeling;
- kennismaken met statistisch onderzoek, populatie en representatieve steekproef

Opgave 6.1

In deze frequentietabel zie je de resultaten van tellingen van het aantal eieren dat dagelijks in een hok vol kippen werd geraapt.

aantal eieren	35	36	37	38	39	40	41	42	43	44
frequentie	1	2	3	6	12	15	27	35	23	12

Tabel 6.1

Deze gegevens kun je samenvatten met behulp van een drietal centrummaten.

- Welk aantal eieren per dag is de modus?
- Welk aantal eieren per dag is de mediaan?
- Bereken het gemiddelde aantal eieren per dag.

Opgave 6.2

In deze frequentietabel zie je de schoenmaten van 124 volwassen mannen.

aantal eieren	39	40	41	42	43	44
frequentie	12	15	27	35	23	12

Tabel 6.2

Bij deze gegevens kun je behalve de drie centrummaten ook een paar spreidingsmaten vaststellen.

- Bepaal de spreidingsbreedte.
- Bepaal de mediaan en de kwartielafstand.
- Teken een bijpassend boxplot.

Opgave 6.3

De lengtes in cm van 60 leerlingen zijn in klassen ingedeeld. Je kunt nu het gemiddelde alleen nog schatten want je weet de ruwe data niet.

- a Bepaal de modale klasse.
- b Hoe schat je het gemiddelde? Bereken deze schatting.

lengteklasse	frequentie
150– < 160	3
160– < 170	17
170– < 180	34
180– < 190	4
190– < 200	2
totaal	60

Tabel 6.3

Opgave 6.4

Gebruik de frequentietabel van **Opgave 6.3**. Door de mediaan en de kwartielen te schatten kun je een boxplot maken.

- a Schat de mediaan.
- b Schat de kwartielen en teken een bijpassend boxplot.

Opgave 6.5

Bij statistisch onderzoek wil je een uitspraak doen over een grote groep (de populatie) op grond van de resultaten van een steekproef uit die groep.

- a Waar moet een goede steekproef aan voldoen?
- b Welke twee soorten statistisch onderzoek ken je op dit moment? Geef van beide een voorbeeld.

Testen

★ Opgave 6.6

Conciërge Martin heeft elke schooldag van drie brugklassen het aantal telaarcomers genoteerd.

aantal leerlingen te laat	0	1	2	3	4	5	6	7
frequentie	5	6	6	8	4	2	3	6

Tabel 6.4

- a Bepaal de modus en de mediaan.
- b Hoeveel weken hield Martin het te laat komen bij?
- c Bereken het gemiddelde in één decimaal nauwkeurig.

★ Opgave 6.7

Tijdens een milieuonderzoek noteerde men in een eikenbos van 30 bomen de omtrek in centimeters. De omtrek is telkens gemeten op een hoogte van 1 m boven de grond.

125	126	128	129	131	131	135	135	135	135	135	138	139	142	142
142	142	143	150	150	151	151	151	155	160	161	165	165	165	166

Tabel 6.5

- a Bepaal de spreidingsbreedte.
- b Bepaal de mediaan.
- c Bepaal beide kwartielen en bereken de kwartielafstand.
- d Teken de bijbehorende boxplot.

★ **Opgave 6.8**

Hier zie je nog een keer de omtrek van 30 bomen in centimeters.

125	126	128	129	131	131	135	135	135	135	135	138	139	142	142
142	142	143	150	150	151	151	151	155	160	161	165	165	165	166

Tabel 6.6

- a Maak bij deze waarnemingsgetallen een frequentietabel. Neem een klassenbreedte van 10 cm en als ondergrens van de eerste klasse 120 cm.
- b Maak een bijpassend staafdiagram.
- c Van hoeveel procent van deze bomen is de omtrek groter dan 150 cm?
- d Schat hoeveel procent van deze bomen een omtrek heeft die groter is dan de gemiddelde omtrek.
- e Je kunt de gemiddelde omtrek ook schatten vanuit de frequentieverdeling. Welke waarde vind je dan voor het gemiddelde? En waarom is dat hoger dan het werkelijke gemiddelde?

★ **Opgave 6.9**

Voor een biologiepracticum zijn op twee velden regenwormen gevangen en is hun lengte gemeten. In de frequentietabel zie je de resultaten. Je gaat de gegevens van beide velden vergelijken.

regenwormen		
lengte (cm)	veld 1	veld 2
0-<3	4	1
3-<6	7	1
6-<9	15	5
9-<12	20	9
12-<15	23	15
15-<18	15	18
18-<21	4	12
21-<24	1	6
24-<27	1	3
	90	70

Figuur 6.1

- a Leg uit waarom het klassenmidden van de eerste klasse 1,5 is.
- b Schat de gemiddelde lengte van de regenwormen op veld 1. Doe hetzelfde voor veld 2.
- c Als je de twee staafdiagrammen bij deze verdelingen wilt vergelijken dan kun je het beste alle frequenties omrekenen naar relatieve frequenties. Waarom is dat? Teken vervolgens beide staafdiagrammen in één figuur.
- d Hoeveel procent van de regenwormen op veld 1 is 15 cm of langer? En op veld 2?

★ **Opgave 6.10**

Gebruik de gegevens over de regenwormen uit de vorige opgave.

- a Welke klasse is de modale klasse op veld 1? En op veld 2?
- b Maak bij beide velden een boxplot van de verdeling van de lengtes van de regenwormen.
- c Kun je een conclusie trekken? Motiveer je antwoord.

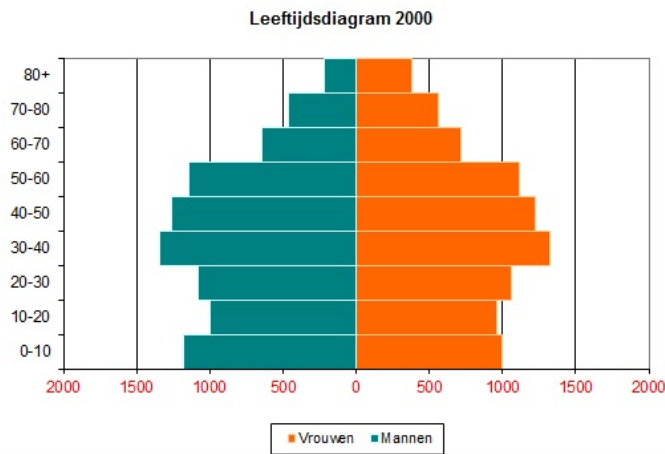
★ **Opgave 6.11**

Stel je voor dat je wilt weten hoeveel procent van de scholieren in het voortgezet onderwijs in Nederland met de fiets naar school gaat. Je gaat dit uitzoeken door middel van een steekproef.

- a Hoe ga je die steekproef samenstellen? Noem minstens twee punten waar je rekening mee moet houden.
- b Zou je dit onderzoek via het internet kunnen uitvoeren? Is dat een goede aanpak?
- c Kun je een betere aanpak verzinnen? Motiveer je antwoord.

Toepassen

Hier zie je een zogenaamde **bevolkingspiramide**, een leeftijdsdiagram. Daarin is de bevolkingsopbouw van Nederland in het jaar 2000 weergegeven. Het lijkt wel een staafdiagram, maar dan op zijn kant.



Figuur 6.2

★★ Opgave 6.12: Bevolkingspiramide in 2000

In **Toepassen** zie je een bevolkingspiramide, een leeftijdsdiagram van de Nederlandse bevolking in het jaar 2000. Alle aantallen zijn duizendtallen.

Je kunt er bijvoorbeeld in aflezen dat de groep 0-10-jarigen in 2000 uit ongeveer 1.200.000 jongens en ongeveer 1.000.000 meisjes bestond.

- Om wat voor soort diagrammen gaat het hier?
- Hoeveel mannen en vrouwen van 10– < 20 waren er in Nederland in 2000?
- In de jaren na de Tweede Wereldoorlog werden er nogal veel kinderen geboren. Dat werd de ‘baby-boom’ genoemd. Hoe vind je dat in de bevolkingspiramide van 2000 terug?
- Wat was in 2000 de modale leeftijdscategorie?
- Schat de gemiddelde leeftijd van de Nederlanders in 2000.
- Het woord ‘bevolkingspiramide’ is wat verouderd. Waarom is dat woord niet meer zo van toepassing als vroeger, denk je?

★★ Opgave 6.13: Bevolkingspiramide vervolg

Bekijk de bevolkingspiramide van de voorgaande opgave nog eens.

- Schat de mediaan van de leeftijden van de Nederlandse mannen in 2000. Doe hetzelfde voor de vrouwen.
- Teken boxplots voor zowel de mannen als de vrouwen in één figuur. Kun je een conclusie trekken? Via de website [CBS: bevolkingspiramide](#) kun je zien hoe de bevolkingsopbouw in Nederland voor dit jaar er uitziet.
- Beschrijf het verschil van de bevolkingsopbouw in Nederland in dit jaar met die in 2000.

Leerdoelentabel

In het achter de opgave kun je aangeven hoe je de opgave hebt gemaakt:

✓ goed gemaakt — **S** wel begrepen maar een slordige fout gemaakt — **H** hulp nodig gehad — **G** samen met groepje goed gemaakt — **X** fout gemaakt en niet goed begrepen — **N** niet bekeken

1	Centrummaten	★	★★	★★★
	Van een hoeveelheid gegevens de modus, de mediaan en het gemiddelde berekenen.	1.1 <input type="checkbox"/> 1.2 <input type="checkbox"/> 1.3 <input type="checkbox"/> 1.4 <input type="checkbox"/> T6.6 <input type="checkbox"/>	1.6 <input type="checkbox"/> 1.8 <input type="checkbox"/> 1.9 <input type="checkbox"/> T6.12 <input type="checkbox"/> T6.13 <input type="checkbox"/>	1.7 <input type="checkbox"/>
	Modus, mediaan en gemiddelde interpreteren als centrummaten van die gegevens.	1.1 <input type="checkbox"/> 1.2 <input type="checkbox"/> 1.3 <input type="checkbox"/> 1.4 <input type="checkbox"/> 1.5 <input type="checkbox"/> T6.6 <input type="checkbox"/>	1.6 <input type="checkbox"/> 1.8 <input type="checkbox"/> 1.9 <input type="checkbox"/> T6.12 <input type="checkbox"/> T6.13 <input type="checkbox"/>	1.7 <input type="checkbox"/>
2	Spreidingsmaten	★	★★	★★★
	De spreidingsbreedte en de kwartielafstand van een serie gegevens berekenen en interpreteren als spreidingsmaten.	2.1 <input type="checkbox"/> 2.2 <input type="checkbox"/> 2.3 <input type="checkbox"/> 2.4 <input type="checkbox"/> 2.5 <input type="checkbox"/> T6.7 <input type="checkbox"/>	2.7 <input type="checkbox"/> 2.8 <input type="checkbox"/>	2.6 <input type="checkbox"/>
	Een serie gegevens samenvatten in een boxplot.	2.1 <input type="checkbox"/> 2.2 <input type="checkbox"/> 2.3 <input type="checkbox"/> 2.5 <input type="checkbox"/> T6.7 <input type="checkbox"/>	2.7 <input type="checkbox"/> 2.8 <input type="checkbox"/>	2.6 <input type="checkbox"/>
3	Klassenindeling	★	★★	★★★
	Een frequentietabel met klassenindeling en geschikte klassenbreedte maken bij een verzameling ruwe data.	3.1 <input type="checkbox"/> 3.2 <input type="checkbox"/> 3.3 <input type="checkbox"/> 3.4 <input type="checkbox"/> 3.5 <input type="checkbox"/> 3.6 <input type="checkbox"/> T6.8 <input type="checkbox"/>	3.7 <input type="checkbox"/> 3.8 <input type="checkbox"/> T6.13 <input type="checkbox"/>	
	Bij zo'n klassenindeling diagrammen maken.	3.5 <input type="checkbox"/> T6.8 <input type="checkbox"/>	3.7 <input type="checkbox"/> T6.12 <input type="checkbox"/> T6.13 <input type="checkbox"/> T6.13 <input type="checkbox"/>	
	Bij een klassenindeling het gemiddelde schatten.	3.2 <input type="checkbox"/> 3.3 <input type="checkbox"/> 3.4 <input type="checkbox"/> 3.5 <input type="checkbox"/> 3.6 <input type="checkbox"/> T6.8 <input type="checkbox"/>	3.8 <input type="checkbox"/> T6.12 <input type="checkbox"/> T6.13 <input type="checkbox"/>	
4	Schattingen	★	★★	★★★
	Bij een klassenindeling centrummaten en spreidingsmaten schatten.	4.1 <input type="checkbox"/> 4.2 <input type="checkbox"/> 4.3 <input type="checkbox"/> 4.4 <input type="checkbox"/> 4.6 <input type="checkbox"/> T6.9 <input type="checkbox"/> T6.10 <input type="checkbox"/>	4.7 <input type="checkbox"/> 4.8 <input type="checkbox"/>	
	Bij een klassenindeling een boxplot maken.	4.1 <input type="checkbox"/> 4.4 <input type="checkbox"/> 4.5 <input type="checkbox"/> T6.10 <input type="checkbox"/>	4.7 <input type="checkbox"/> 4.8 <input type="checkbox"/>	
5	Statistische uitspraken	★	★★	★★★
	Een statistische uitspraak herkennen.	5.3 <input type="checkbox"/>	5.4 <input type="checkbox"/> 5.5 <input type="checkbox"/>	
	Bij een statistisch onderzoek de begrippen populatie en steekproef herkennen.	5.1 <input type="checkbox"/> 5.2 <input type="checkbox"/> 5.3 <input type="checkbox"/> T6.11 <input type="checkbox"/>	5.5 <input type="checkbox"/>	
	Aangeven of een steekproef representatief is.	5.1 <input type="checkbox"/> 5.2 <input type="checkbox"/> 5.3 <input type="checkbox"/> T6.11 <input type="checkbox"/>	5.5 <input type="checkbox"/>	
	Een eenvoudig statistisch onderzoek opzetten.	5.3 <input type="checkbox"/>	5.4 <input type="checkbox"/> 5.5 <input type="checkbox"/>	

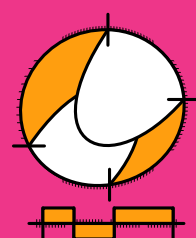
Het lesmateriaal in deze reader is gebaseerd op het materiaal dat ook op de Math4All website staat.

De reader is gegenereerd met de Math4All maatwerkdienst. De inhoud en de volgorde van de onderwerpen in deze reader zijn gekozen door docenten van het ConTeXt College.

Stichting Math4All



www.math4all.nl



Informatieblad bij Opdracht 1.1

Je ziet hier de rapportcijfers voor het vak Engels in B1H.

4	5	7	5	10	9	8	4	6	6	7	7	8	3	7
6	6	7	7	9	7	6	7	6	6	8	7	7	8	

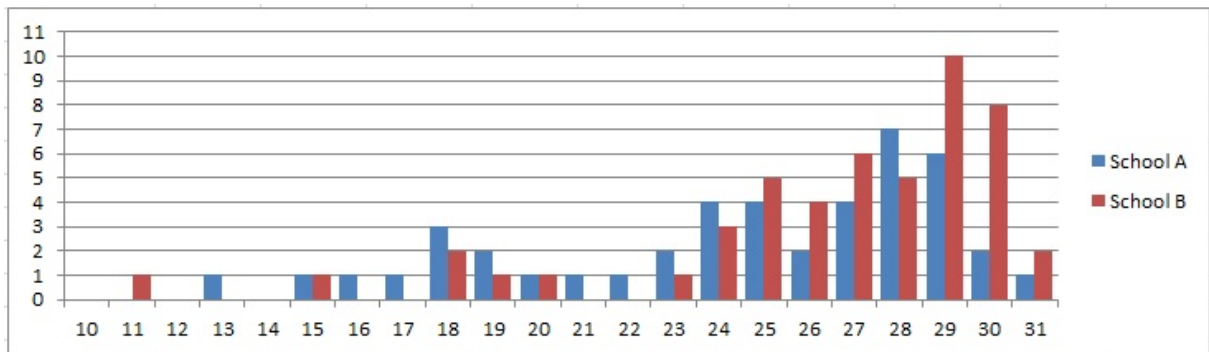
Informatieblad bij Opdracht 1.2

Je ziet hier de frequentietabellen van de klassen B2A en B2C van hun rapportcijfers voor het vak wiskunde.

	klas B2A	klas B2C
cijfer	frequentie	frequentie
4	1	0
5	4	4
6	9	8
7	11	6
8	3	4
9	1	2
totaal	29	24

Informatieblad bij Opdracht 1.3

Op twee scholen voor voortgezet onderwijs zijn de aantallen leerlingen per klas geteld. Je ziet in dit staafdiagram het resultaat voor school A en school B.



Informatieblad bij Opdracht 2.1

Je ziet hier de rapportcijfers voor het vak Engels in B1H.

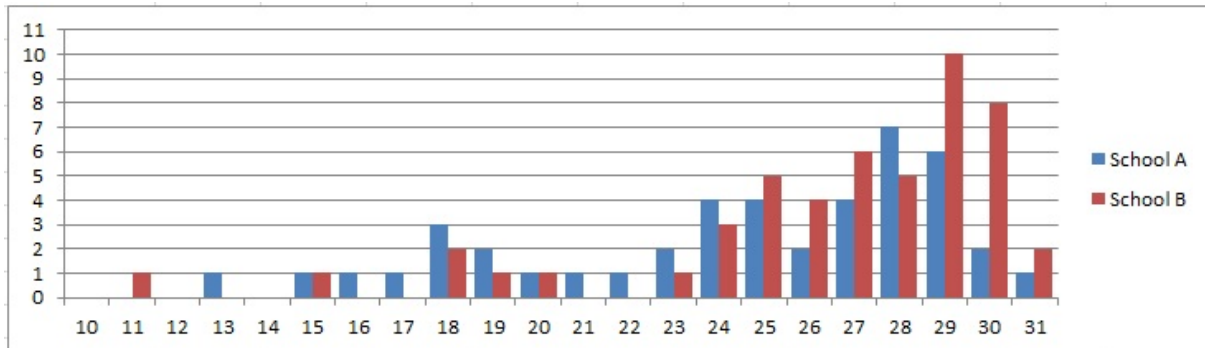
cijfer	frequentie
3	1
4	2
5	2
6	7
7	10
8	4
9	2
10	1
totaal	29

Informatieblad bij Opdracht 2.2

		wi						
	B2d	4	5	6	7	8	9	10
sc	4		1					
	5		2	3	1			
	6	1	4	2	1	1		
	7			1	3	2	1	
	8					2	1	
	9							1

Informatieblad bij Opdracht 2.3

Op twee scholen voor voortgezet onderwijs zijn de aantallen leerlingen per klas geteld. Je ziet in dit staafdiagram het resultaat voor school A en school B.



Informatieblad bij Opdracht 3.1

Dit zijn dertig rapportcijfers van klas 2A op één decimaal nauwkeurig.

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

Informatieblad bij Opdracht 3.2

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

Informatieblad bij Opdracht 3.3

Deze tabel laat de verdeling zien van de lengtes van 90 meisjes in een vierde klas. De gebruikte klassenindeling heeft een klassenbreedte van 5.

lengteklasse	klassenmidden	frequentie
150– < 155	152,5	1
155– < 160	157,5	5
160– < 165	162,5	10
165– < 170	167,5	16
170– < 175	172,5	22
175– < 180	177,5	15
180– < 185	182,5	11
185– < 190	187,5	6
190– < 195	192,5	3
195– < 200	197,5	1
	totaal	90

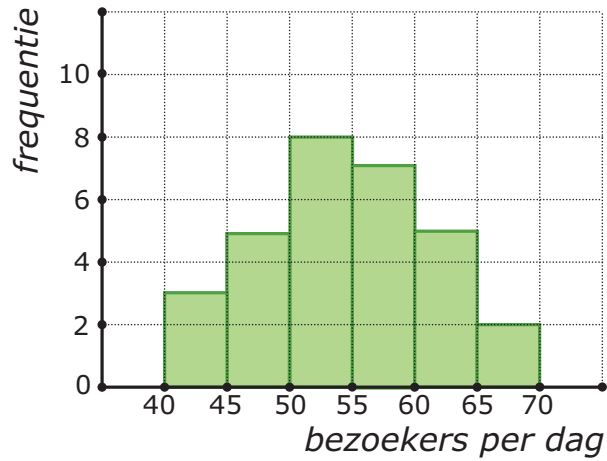
Informatieblad bij Opdracht 4.1

Dit is een klassenindeling van de rapportcijfers van klas 2A.

klasse	klassenmidden	frequentie
3,5– < 4,5	4	2
4,5– < 5,5	5	2
5,5– < 6,5	6	7
6,5– < 7,5	7	11
7,5– < 8,5	8	5
8,5– < 9,5	9	3
totaal		30

Informatieblad bij Opdracht 4.2

Dit staafdiagram geeft voor de maand juni de aantallen bezoekers van een klein museum.



Informatieblad bij Opdracht 5.1

1. "40% van de leerlingen in klas 2C van school X is een meisje."
2. "40% van de Nederlanders heeft bloedgroep O-positief."
3. "85% van alle jongeren tussen 12 en 18 jaar internet dagelijks."
4. "Ik heb zojuist in 360 worpen met twee dobbelstenen precies 60 keer 7 ogen gegooid."
5. "Als je met twee dobbelstenen gooit zul je in 16,7% van alle worpen 7 ogen gooien."
6. "Jongens kunnen verder gooien dan meisjes."

