



2

Statistiek

Wat leer je?

- Hoe de statistische cyclus wordt gebruikt bij het uitvoeren van een onderzoek.
- Wat een representatieve steekproef is.
- Welke soorten statistische variabelen er zijn en hoe je ze onderscheidt.
- Gebruiken van verschillende centrum- en spreidingsmaten.
- Werken met verschillende diagrammen waarin statistische data gepresenteerd staan.
- Verschillende soorten verdelingen, waaronder de normale verdeling, herkennen en gebruiken.
- Wat een betrouwbaarheidsinterval is.



Beginopdracht Gezondheid en welzijn van jongeren

In 2018 publiceerden de Universiteit Utrecht, het Trimbos-instituut en het Sociaal en Cultureel Planbureau gezamenlijk het rapport Gezondheid en welzijn van jongeren in Nederland (HBSC 2017). Het internationale HBSC-onderzoek bestaat sinds 1983 en is een van de langstlopende onderzoeken naar de jeugd in Europa. Het onderzoek wordt elke vier jaar opnieuw uitgevoerd en de resultaten in onderstaande tabel gaan over de stand van zaken bij 6718 Nederlandse middelbare scholieren in 2017.

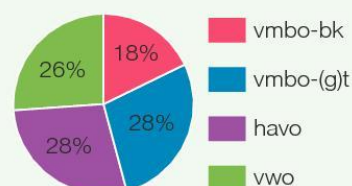
	geslacht		schoolniveau			
	jongen	meisje	vmbo-bk	vmbo-(g)t	havo	vwo
gedurende de hele dag contact via sociale media	890	1214	481	610	525	487
problematisch gebruik sociale media	217	282	141	166	114	80
minstens 24 uur per week gamen	464	59	125	198	99	67
problematisch gamegedrag	244	26	58	91	56	59
minstens 2× per week sport	2870	2405	878	1458	1376	1563
BMI overgewicht	385	279	213	233	121	105
voelt zich te dik	804	1431	431	680	586	575
totaal	3437	3281	1205	1906	1703	1904

Gebruik de gegevens uit de tabel om de volgende opdrachten te maken.

- Trek een conclusie over *gedurende de hele dag contact via sociale media* waarbij je onderscheid maakt in geslacht. Doe dat ook voor *minstens 24 uur per week gamen*. Wat valt je op?
- Trek een conclusie over *BMI overgewicht* waarbij je onderscheid maakt in schoolniveau. Doe dat ook voor *voelt zich te dik*. Wat valt je op?
- In het cirkeldiagram hiernaast zie je de verdeling van leerlingen naar schoolniveau in het schooljaar 2017/2018.

Is de verdeling van leerlingen naar schoolniveau in het HBSC-onderzoek in overeenstemming met de gegevens in het cirkeldiagram? Licht toe.

VERDELING VAN LEERLINGEN NAAR SCHOOLNIVEAU 2017/2018



Voorkennis Procenten en frequenties

Theorie A Procentberekeningen

Bereken $p\%$ van A .

18% betekent 18 van de 100, dus $18\% = \frac{18}{100} = 0,18$.

18% van 620 is $0,18 \times 620 = 111,6$.

Hoeveel is $p\%$ van A ? Antwoord: $\frac{p}{100} \times A$.

Bereken hoeveel procent A van B is.

In een klas zitten 28 leerlingen van wie er zes zijn blijven zitten.

Dat is $\frac{6}{28} \times 100\% \approx 21,4\%$.

Rond percentages af op één decimaal.

Hoeveel procent is A van B ? Antwoord: $\frac{A}{B} \times 100\%$.

Procentuele verandering

Op een school zitten 1315 leerlingen. Vorig schooljaar waren dat er 1248.

De absolute verandering is $1315 - 1248 = 67$.

De relatieve verandering in procenten is $\frac{67}{1248} \times 100\% \approx 5,4\%$.

De relatieve verandering in procenten heet de procentuele verandering.

Je ziet procentuele verandering = $\frac{\text{absolute verandering}}{\text{OUD}} \times 100\%$.

Absolute verandering = NIEUW – OUD

Procentuele verandering = $\frac{\text{NIEUW} - \text{OUD}}{\text{OUD}} \times 100\%$

Bij een procentuele verandering van -12% is sprake van een procentuele afname van 12% .

Bereken hoeveel je krijgt bij een toename of afname van $p\%$.

Een hoeveelheid van 250 neemt met 8% toe.

De nieuwe hoeveelheid is $1,08 \times 250 = 270$.

..... $100\% + 8\% = 108\%$, dus $\times 1,08$

Een hoeveelheid van 1800 neemt met 11% af.

De nieuwe hoeveelheid is $0,89 \times 1800 = 1602$.

..... $100\% - 11\% = 89\%$, dus $\times 0,89$

Hoeveel krijg je bij een verandering van $p\%$?

Je krijgt NIEUW = $\left(1 + \frac{p}{100}\right) \times \text{OUD}$.

Bij een afname van $p\%$ neem je p negatief in bovenstaande formule.

Bereken het totaal.

Een hoeveelheid van 298 is 58% van het totaal.

Dus $0,58 \times \text{totaal} = 298$. Dan is $\text{totaal} = \frac{298}{0,58} \approx 514$.

$p\%$ van het totaal is A . Hoeveel is het totaal?

Antwoord: $\text{totaal} = \frac{A}{0,01p}$.

Geef het antwoord
in dezelfde
nauwkeurigheid als het
gegeven.

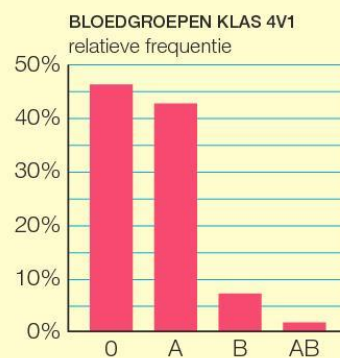
- 1 a Hoeveel procent is 57 minder dan 121?
b Hoeveel is 0,3% van 8700?
c Een hoeveelheid van 3,2 miljoen neemt 8,7% af. Hoeveel krijg je?
d Een hoeveelheid van 8,3 miljoen is 12,6% van het totaal.
Hoeveel is het totaal?
- 2 In 1950 waren er 2521 miljoen mensen op de wereld. Hiervan woonden er 1402 miljoen in Azië, 547 miljoen in Europa en 221 miljoen in Afrika. In 2017 was de wereldbevolking toegenomen tot 7550 miljoen. De procentuele toename was het grootst in Afrika met een toename van maar liefst 468%. In 2017 waren er in Azië 4504 miljoen inwoners en in Europa 742 miljoen.
a Bereken de procentuele toename van de wereldbevolking in de periode 1950-2017.
b Bereken in miljoenen nauwkeurig het aantal inwoners van Afrika in 2017.
c Onderzoek met een berekening of de procentuele toename van de bevolking in Azië in de periode 1950-2017 inderdaad minder is dan de procentuele toename in Afrika in deze periode.
d Bereken zowel voor het jaar 1950 als voor het jaar 2017 hoeveel procent van de wereldbevolking in Europa woonde.

Theorie B Frequentietabellen en staafdiagrammen

In de frequentietabel hieronder staan de (absolute) frequenties bij de bloedgroepen van de leerlingen in klas 4V1. De totale frequentie is $13 + 12 + 2 + 1 = 28$.

BLOEDGROEPEN KLAS 4V1

bloedgroep	frequentie
0	13
A	12
B	2
AB	1



figuur 2.1

In figuur 2.1 is bij de tabel een staafdiagram met relatieve frequenties getekend. Zo is de relatieve frequentie van bloedgroep 0 gelijk aan

$$\frac{13}{28} \times 100\% \approx 46,4\%.$$

- 3 In de frequentieverdeling hiernaast zie je waar de leerlingen van klas 4VA wonen.

a Bereken de relatieve frequentie bij Schalkhaar.

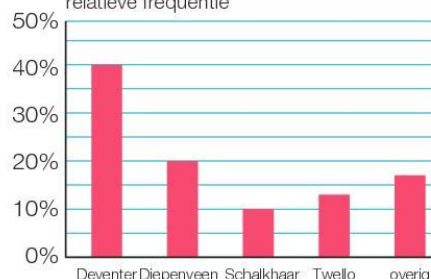
WOONPLAATSEN KLAS 4VA

woonplaats	frequentie
Deventer	11
Diepenveen	3
Schalkhaar	3
Twello	2
overig	5

In figuur 2.2 zie je een staafdiagram met de verdeling van de woonplaatsen bij de leerlingen van klas 4VB. Er zijn 6 leerlingen in klas 4VB die in Diepenveen wonen.

b Hoeveel leerlingen zitten er in klas 4VB?

WOONPLAATSEN KLAS 4VB
relatieve frequentie



figuur 2.2

- 4 In figuur 2.3 zie je een samengesteld staafdiagram, ook wel stapeldiagram genoemd.

Het diagram komt uit het rapport Gezondheid en welzijn van jongeren in Nederland in 2017.

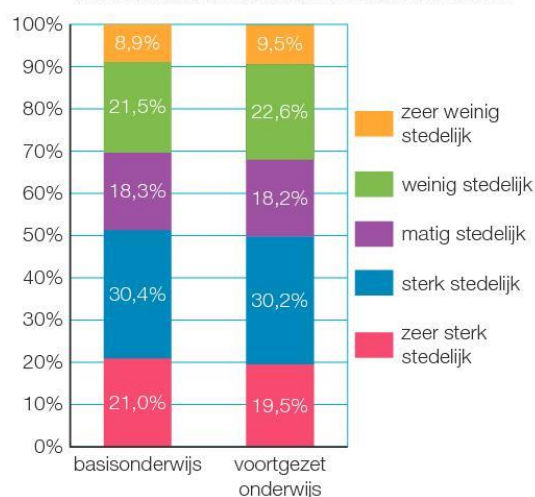
Je ziet dat in het basisonderwijs het aandeel deelnemers uit sterk stedelijk gebied gelijk is aan 30,4%.

a Er waren in het voortgezet onderwijs 1310 deelnemers uit zeer sterk stedelijk gebied.

Hoeveel deelnemers waren er in totaal in het voortgezet onderwijs?

b Er waren in het basisonderwijs 51 deelnemers meer uit weinig stedelijk gebied dan uit matig stedelijk gebied. Hoeveel deelnemers waren er in totaal in het basisonderwijs?

VERDELING DEELNEMERS NAAR STEDELIJK GEBIED



figuur 2.3

2.1 De statistische cyclus

01



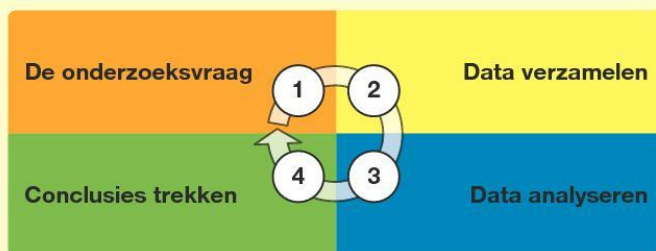
Daan en Emma hebben een onderzoek gedaan. Hieronder staan vijf activiteiten die ze hebben uitgevoerd bij dit onderzoek.

Welke volgorde is logisch?

1. Het interpreteren van alle verzamelde informatie om te kijken of de onderzoeksvraag hiermee beantwoord kan worden.
2. Een overzicht maken van alle aspecten die bij het onderwerp komen kijken, om zo tot de onderzoeksvraag te komen.
3. Het beantwoorden van de onderzoeksvraag op basis van de uitkomsten van de verzamelde enquêtes en gevonden informatie.
4. Vaststellen welke kenmerken onderzocht moeten worden om de onderzoeksvraag te kunnen beantwoorden.
5. Het afnemen van enquêtes en opzoeken van informatie over het onderwerp.

Theorie A Fasen in de statistische cyclus

Bij een statistisch onderzoek probeert men antwoord op een vraag te vinden door het verzamelen, verwerken en interpreteren van gegevens. De **statistiek** houdt zich bezig met methoden en technieken voor het verzamelen, verwerken, analyseren en presenteren van gegevens. De verzamelde gegevens worden **data** genoemd. Deze data staan centraal in de **statistische cyclus**: het proces van opstellen van een onderzoeksvraag, verzamelen en analyseren van data en hieruit conclusies trekken. Het woord cyclus wijst erop dat het doorlopen van deze stappen vaak nieuwe (vervolg)vragen oplevert die verder onderzocht kunnen worden.



figuur 2.4 De statistische cyclus

In dit hoofdstuk staan de vier fasen van de statistische cyclus centraal. Wanneer je zelf een onderzoek doet (bijvoorbeeld voor je profielwerkstuk) helpt de statistische cyclus bij het in goede banen leiden van het proces.



2

Het onderzoek Gezondheid en welzijn van jongeren in Nederland (HBSC 2017) richt zich onder andere op schoolbeleving. Een van de onderzochte aspecten is het ervaren van druk door schoolwerk. De onderzoekers hebben veel stappen gezet voordat ze conclusies over dit onderwerp konden trekken. Hieronder staat een aantal van deze stappen. Zet ze in de juiste volgorde en geef aan bij welke fase van de statistische cyclus ze horen.

- A In de enquête is een vraag opgenomen over in hoeverre druk wordt ervaren door schoolwerk. Leerlingen kunnen kiezen uit ‘helemaal niet’, ‘een beetje’, ‘nogal veel’ en ‘veel’.
- B De onderzoekers concluderen dat het ervaren van druk door schoolwerk tussen 12 en 14 jaar aanzienlijk toeneemt.
- C Er is overleg geweest over welke aspecten een rol spelen bij de schoolbeleving. Een van deze aspecten bleek het ervaren van druk door schoolwerk. De onderzoekers wilden weten in hoeverre leerlingen druk ervaren door schoolwerk.
- D De onderzoekers hebben besloten om schoolwerk niet nader te specificeren, zodat alle taken die met school te maken hebben worden meegewogen.
- E Uit de ingevulde enquêtes blijkt dat van de 12-, 13- en 14-jarige middelbare scholieren respectievelijk 24,5%, 33,7% en 41,0% (nogal) veel druk ervaart door schoolwerk.
- F Men adviseert middelbare scholen om vanaf de eerste klas aandacht te besteden aan het omgaan met werkdruk. Ook wordt scholen gevraagd om werkdruk van leerlingen in de onderbouw in de gaten te houden middels gesprekken met de mentor.
- G Er is besloten om de resultaten uit te splitsen naar geslacht, leeftijd en schoolniveau.

INFORMATIEF

Gedeelte van het voorwoord HBSC 2017

‘De jeugd van tegenwoordig’ staat volop in de belangstelling. Worden jongeren in Nederland gezonder of minder gezond? Moeten we ons nog zorgen maken over hun alcoholgebruik? Zitten jongeren niet te veel op hun mobiel? Hoe heeft hun relatie met klasgenoten, vrienden en ouders zich ontwikkeld in de laatste jaren? Voelen jongeren in Nederland zich steeds meer onder druk staan door hun schoolwerk? Hoeveel psychische problemen ervaren zij? Maar ook: met welke jongeren gaat het goed en met welke juist niet? Om een antwoord te geven op deze, en vele andere vragen, vond in het najaar van 2017 voor de vijfde keer het Nederlandse Health

Behaviour in School-aged Children (HBSC) onderzoek plaats. Medewerkers van de Universiteit Utrecht, het Trimbos-instituut en het Sociaal en Cultureel Planbureau (SCP) voerden het onderzoek in Nederland uit, gefinancierd door het ministerie van Volksgezondheid, Welzijn en Sport (VWS). Dit onderzoek vindt elke vier jaar plaats onder een landelijk representatieve groep leerlingen in een groot aantal landen in Europa, Noord-Amerika en Israël. De onderzoeksgegevens maken het mogelijk om een goed beeld te krijgen van de gezondheid en het welzijn van Nederlandse scholieren in de leeftijd van 11 tot en met 16 jaar.

INFORMATIEF

Kijkonderzoek

Voor veel Nederlanders is televisiekijken een belangrijke tijdsbesteding. Per dag kijken we in Nederland per persoon ongeveer 3 uur televisie. De kijkcijfers zijn voor televisiemakers en zenders een graadmeter voor succes.

Onderzoeksbureau GfK voert kijkcijferonderzoek uit voor de Stichting KijkOnderzoek (SKO). Het kijkonderzoek laat zien hoeveel mensen

er kijken naar zenders en programma's. Met behulp van deze gegevens kunnen inhoud en tijdstip van programma's afgestemd worden op wat de kijker belangrijk vindt. Het succes van een programma telt bovendien mee bij het bepalen van het tarief voor een reclamespot in de omliggende reclameblokken.

Hieronder staan vier fragmenten uit het kijkonderzoek.

Geef voor elk van de fragmenten aan bij welke fase in de statistische cyclus dit het beste past.

A

Nadat de metergegevens van alle panelhuishoudens naar GfK zijn gestuurd, wordt vastgesteld naar welke zenders is gekeken. Dit gebeurt door middel van Enhanced Audio Matching; bij deze techniek wordt gebruikgemaakt van het matchen van audio samples.

B

SKO stelt dagelijks de standaardresultaten kijkdichtheid, marktaandeel en de absolute aantallen kijkers (KijkTotaal en Uitzenddag) ter beschikking. Deze informatie komt op de website en is bedoeld voor het algemene publiek. De resultaten die SKO op de website publiceert, zijn onder voorwaarden te gebruiken en de bronvermelding is verplicht.

C

Op basis van de achtergrondkenmerken van de panelleden zijn verschillende vragen te onderzoeken: hoeveel 13 t/m 19-jarigen kijken naar een bepaald programma, hoeveel autobezitters hebben de commercial voor een nieuwe personenwagen gezien, kijken alleen sportliefhebbers naar sportprogramma's? Het is belangrijk de vragen scherp te formuleren en af te bakenen voor er tot verder onderzoek wordt overgegaan.

D

Elk huishouden dat meedoet aan het kijkonderzoek krijgt op alle televisietoestellen in huis een 'kijkmeter'. De kijkmeter kan automatisch registreren dat een toestel aanstaat en slaat deze informatie op. De kijkmeter legt vast wat er in het televisietoestel gebeurt, maar kan niet 'zien' wat zich in de kamer afspeelt en wie er naar de televisie kijken. Daarom moeten alle personen van 3 jaar en ouder die kijken zich aanmelden.

bron: SKO

Hieronder staan enkele onderzoeksvragen waar iets mis mee is.

Geef bij elke vraag aan waarom het geen goede onderzoeksvraag is.

- a Hoeveel jongeren eten gezond?
- b Vertonen jongeren die veel gamen vaak probleemgedrag en hoe kunnen we dat verbeteren?
- c Hoe komt het dat jongeren die veel gamen minder sociaal zijn?

Theorie B De onderzoeksvraag

Voor elk onderzoek is de onderzoeksvraag, ook wel hoofdvraag genoemd, van cruciaal belang. Een goede onderzoeksvraag helpt je om je onderzoek gedegen uit te voeren en om een overzichtelijk verslag te maken.

Een onderzoeksvraag stelt een verband tussen twee of meer zaken aan de orde. Dit verband kan een *samenhang* aangeven, zoals bij de onderzoeksvraag ‘In welke leeftijdscategorie wordt er door jongeren het meest gesport?’.

Het verband kan bovendien *causaal* zijn. Je hebt met een **causaal verband** oftewel **oorzakelijk verband** te maken als een gebeurtenis het directe gevolg is van een andere gebeurtenis.

Een voorbeeld van een onderzoeksvraag met een causaal verband is ‘Hoe beïnvloedt slaapgedrag het concentratievermogen tijdens de les?’.

INFORMATIEF

Causaliteit in de rechtspraak: het etalageruit-arrest

Een verdachte reed met een te hoge snelheid in een auto door een winkelstraat in Den Haag. Door de regen en de natte tramrails raakte hij in een slip en botste tegen een verkeerszuil. Daarbij kantelde de auto en schoof, op het dak liggend, door naar het trottoir. Daar raakte de auto een voetganger die daardoor tegen een etalageruit kwam. Door het vallende glas van de gebroken etalageruit raakte deze man ernstig gewond, terwijl een vrouw dodelijk gewond raakte.

De vraag was: is er een causaal verband tussen het verkeersongeval en het feit dat de vrouw overleed? De verdediging stelde dat de dood van de vrouw was veroorzaakt door het vallende glas van de etalageruit en dus niet door de botsing of aanrijding van de verdachte. De Hoge Raad oordeelde echter dat er een oorzakelijk verband was tussen de botsing van de auto en de dood van de vrouw door het vallende glas.

Het opstellen van een onderzoeksvraag gaat in vier stappen.

1 Het onderzoeksterrein *verkennen*.

Om een goede vraag te stellen, dien je voldoende te weten van het onderwerp. Het kan zijn dat je al langere tijd geïnteresseerd bent in het onderwerp en daar al wat van weet. Anders is het verstandig je globaal in te lezen.

2 Het onderzoeksterrein *afbakenen*.

Beperk je bij het onderzoek tot iets dat je overziet. Maak je het onderzoeksterrein te breed, dan bestaat de kans dat je de grote hoeveelheid informatie die op je afkomt niet kunt overzien, of dat je te veel tijd kwijt bent met het onderzoek.

3 De *eerste versie* van de onderzoeksvraag *formuleren*.

Nadat je het onderzoeksterrein voldoende hebt afgebakend, blijven er nog verschillende soorten vragen over die je kunt stellen. Wil je een vraag stellen waarmee je het verschijnsel dat je gaat onderzoeken *beschrijft*, of wil je meer de kant op van een *verklaring* of van een *oordeel* of een *advies*?

4 De onderzoeksvraag *toetsen*.

Een onderzoeksvraag moet aan de volgende eisen voldoen.

- De onderzoeksvraag moet *scherp geformuleerd* zijn.
Er mogen dus geen vage woorden in zitten.
De vraag in opgave 4a kan dus beter worden geformuleerd als ‘Welk percentage van de middelbare scholieren eet dagelijks volgens de richtlijnen van het voedingscentrum?’.
- De onderzoeksvraag moet *enkelvoudig* zijn.
Aan het woordje ‘en’ in de vraag van opgave 4b kun je zien dat het in feite om twee onderzoeken gaat. Een betere vraag zou zijn ‘Welke invloed heeft overmatig gamen op het ontwikkelen van probleemgedrag van jongeren?’.
- De onderzoeksvraag moet *nieuw* zijn.
Het heeft geen zin een onderzoek te doen naar iets dat algemeen bekend is. Een vraag als ‘Heeft op jonge leeftijd beginnen met roken een negatief effect op de gezondheid?’ heeft geen zin, omdat het antwoord al bekend is.
- De onderzoeksvraag mag *niet uitgaan van foute veronderstellingen*.
De vraag in opgave 4c gaat uit van de veronderstelling dat jongeren die veel gamen minder sociaal zijn. Een betere vraag zou zijn ‘Welke invloed heeft gamen op het sociale gedrag van jongeren?’.

5 Hieronder staan enkele onderzoeksvragen.



Welke van deze vragen stellen een causaal verband aan de orde?

Licht toe.

- a** Welk effect heeft overmatig gamen op schoolprestaties?
- b** Welk verband is er bij scholieren tussen gebruik van sociale media en gamen?
- c** In hoeverre helpt het verbieden van roken op het schoolplein bij het terugdringen van roken onder scholieren?
- d** Hoe hangt het rookgedrag van scholieren samen met het opleidingsniveau?

6 Karlijn is geïnteresseerd in paardrijden. Ze rijdt al jaren paard en wil onderzoeken of kinderen van rijke ouders vaker paardrijden dan kinderen van ouders die niet rijk zijn.



- a** Zou Karlijn het onderzoeksterrein nog moeten verkennen? Licht toe.
- b** Bedenk een onderzoeksvraag voor Karlijn.
- c** Is de onderzoeksvraag die je hebt bedacht voldoende afgebakend? Licht toe. Verander zo nodig de onderzoeksvraag.
- d** Ga na of de onderzoeksvraag die je hebt bedacht voldoet aan de vier eisen die in de theorie staan vermeld. Licht je antwoorden toe en verander zo nodig de onderzoeksvraag.

**7**

Nadat de onderzoeksvraag is opgesteld en voordat je data gaat verzamelen, is het soms nodig om een aantal *deelvragen* te bedenken. Om de onderzoeksvraag te beantwoorden, moet je eerst een aantal zaken weten. Het beantwoorden van de deelvragen zou moeten leiden tot het beantwoorden van de hoofdvraag.

Zo zouden bij de onderzoeksvraag ‘Welke invloed heeft overmatig gamen op het ontwikkelen van probleemgedrag van jongeren?’ de volgende deelvragen geformuleerd kunnen worden.

1. Welk effect heeft overmatig gamen op schoolprestaties?
2. Welk effect heeft overmatig gamen op sociaal gedrag?

Bedenk bij elk van de onderzoeksvragen hieronder twee deelvragen.

- a** Welk percentage van de middelbare scholieren eet dagelijks volgens de richtlijnen van het voedingscentrum?
- b** Met welk percentage wordt overgewicht bij jongeren verminderd als alle jongeren minstens vijf uur per week intensief zouden sporten?

Zie opgave 6.

- c** Bedenk bij de onderzoeksvraag die je voor Karlijn had bedacht twee deelvragen.

**8**

Bedenk bij twee van de onderzoeksvragen van opgave 5 twee deelvragen.

**9**

In de theorie worden vier soorten onderzoeksvragen genoemd:

- *** A een beschrijvende vraag
- B een verklarende vraag
- C een vraag die uitmondt in een oordeel
- D een vraag die uitmondt in een advies.

Hieronder staan vier onderzoeksvragen die horen bij een onderzoek naar de ontwikkeling van de woningmarkt in Nederland.

- 1 Hoe is de prijsontwikkeling op de Nederlandse woningmarkt van de afgelopen vijf jaar te verklaren?
 - 2 Welk beleid moet er worden gevoerd om een stabiele woningmarkt in Nederland te bereiken?
 - 3 Hoe is de Nederlandse woningmarkt veranderd sinds het jaar 2000?
 - 4 In welke mate voldoet de ontwikkeling op de Nederlandse woningmarkt aan de belofte van de regering?
- a** Zeg bij elk van de genoemde onderzoeksvragen van welke soort (A, B, C of D) de vraag is.
 - b** Bedenk bij twee van de onderzoeksvragen hierboven twee deelvragen.



Voor een kijkonderzoek wordt niet van elke Nederlander onderzocht welke tv-programma's zijn bekeken.

- a Noem een mogelijke reden waarom niet van elke Nederlander het kijkgedrag wordt onderzocht.

Over de groep Nederlanders in het onderzoek schrijft Stichting KijkOnderzoek (SKO) het volgende.

‘Het is voldoende om een representatieve groep te onderzoeken.

Representatief wil zeggen dat de groep net zo is opgebouwd als de bevolking, met een overeenkomstig percentage jongeren, vrouwen, hoger opgeleiden, enzovoort.’

- b Waarom denk je dat SKO wil dat de te onderzoeken groep representatief is?

Theorie C Een steekproef nemen

Een statistisch onderzoek hoeft zich niet alleen te richten op individuele mensen, maar kan ook betrekking hebben op bijvoorbeeld huishoudens, scholen en materiële zaken zoals lampen of pakken melk. De totale groep waarop het onderzoek zich richt heet de **populatie**.

Wil men onderzoeken hoeveel procent van de stemgerechtigde Nederlanders geïnteresseerd is in de politiek, dan wordt de populatie gevormd door de stemgerechtigde Nederlanders. Voor het onderzoek wordt echter niet elke stemgerechtigde Nederlander ondervraagd. Er wordt een **steekproef** genomen.

Het is van belang dat een steekproef een goede weergave is van de totale populatie. We zeggen dat de steekproef **representatief** moet zijn.

Een steekproef is representatief als zij een goede afspiegeling is van de gehele populatie.

Een representatieve steekproef moet in elk geval voldoen aan de volgende eisen.

- **De steekproef is voldoende groot.**
Hoe meer kenmerken je wilt onderzoeken (zoals leeftijd, werk en inkomen), hoe groter de steekproef moet zijn.
- **De steekproef is aselekt.**
Een steekproef is **aselect** als elk element van de populatie een even grote kans heeft om in de steekproef voor te komen.

Om een representatieve steekproef te krijgen bij de maandelijkse peiling naar de politieke voorkeur, worden ongeveer 3000 stemgerechtigde Nederlanders ondervraagd. De personen in de steekproef worden zo gekozen dat alle Nederlandse stemgerechtigden een even grote kans hebben om in de steekproef voor te komen.

11



Omschrijf bij de volgende onderzoeksvragen een populatie.

- a Welke vervolgoopleidingen kiezen de leerlingen van je school?
- b In welk weekblad staan de meeste illustraties?
- c Welk percentage van de inwoners van Appelscha is daar ook geboren?
- d Hoe effectief is een nieuw geneesmiddel tegen hooikoorts?

12

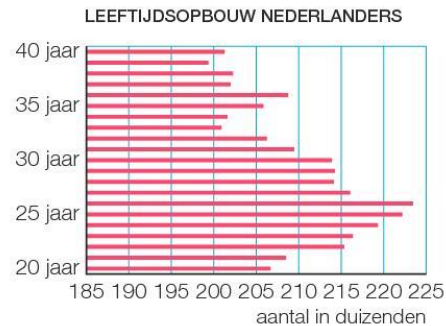


- a Er wordt een onderzoek gedaan naar de fruitconsumptie van Nederlanders van 20 tot en met 40 jaar door van elke leeftijd evenveel personen te ondervragen.

Geef commentaar. Gebruik de figuur hiernaast.

- b Bij een onderzoek worden alleen mannen in de pensioengerechtigde leeftijd ondervraagd. Kun je hier te maken hebben met een aselechte steekproef? Licht toe.

- c Bij een onderzoek worden 100 mensen ondervraagd. Kun je hier te maken hebben met een steekproef die voldoende groot is? Licht toe.



figuur 2.5

A13



Een onderzoeksbureau wil de mening weten van de Nederlander over klimaatverandering.

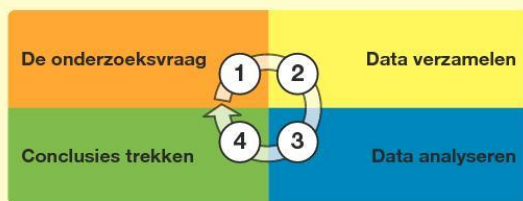
Waarom zijn de volgende steekproeven niet representatief?

- a Op een zaterdagmiddag wordt op een druk kruispunt in de stad een groot aantal voorbijgangers ondervraagd.
- b In de ochtendspits wordt in een file een groot aantal automobilisten ondervraagd.
- c Door loting via postcodenummers wordt uit elke provincie één persoon ondervraagd.
- d Een groot aantal toeristen in het bezoekerscentrum van het natuurgebied de Posbank wordt ondervraagd.
- e Er wordt een vragenlijst afgenomen via de smartphone.

Terugblik

De statistische cyclus

De statistiek houdt zich bezig met methoden en technieken voor het verzamelen, verwerken, analyseren en presenteren van gegevens. De verzamelde gegevens worden data genoemd. Deze data staan centraal in de statistische cyclus: het proces van opstellen van een onderzoeksvraag, verzamelen en analyseren van data en hieruit conclusies trekken. Het woord cyclus wijst erop dat het doorlopen van deze stappen vaak nieuwe (vervolg)vragen oplevert die verder onderzocht kunnen worden.



De onderzoeksvraag

Een statistisch onderzoek begint met een goede onderzoeksvraag. Deze vraag stelt een verband tussen twee of meer zaken aan de orde.

Bij de vraag ‘Op welke leeftijd gamen leerlingen het meest?’ is er sprake van een samenhang tussen leeftijd en gamegedrag.

Bij de vraag ‘Welke invloed heeft gamegedrag op de relatie van een kind met de ouders?’ is er sprake van een causaal (oorzakelijk) verband: het directe gevolg van gamegedrag van het kind wordt onderzocht.

De vier stappen van het opstellen van een onderzoeksvraag zijn:

- 1 Het onderzoeksterrein verkennen.
- 2 Het onderzoeksterrein afbakenen.
- 3 De eerste versie van de onderzoeksvraag formuleren.
- 4 Controleren of de onderzoeksvraag scherp geformuleerd, enkelvoudig en nieuw is en of de vraag niet uitgaat van foute veronderstellingen.

Een steekproef nemen

De totale groep waarop een onderzoek zich richt heet de populatie.

De populatie kan een groep mensen zijn, maar bijvoorbeeld ook geneesmiddelen of elektrische auto's.

Het is vaak niet nodig en ook ondoenlijk om de gehele populatie te onderzoeken. Het nemen van een representatieve steekproef is voldoende.

Een steekproef is representatief als zij een juiste afspiegeling is van de gehele populatie.

Een representatieve steekproef moet in ieder geval voldoende groot en aselekt zijn.

2.2 Data verzamelen

014
□ ⊗ *

Bij een onderzoek wordt van alle sporters die deelnemen aan de Olympische Spelen een aantal kenmerken genoteerd.

- a Bedenk vijf van deze mogelijke kenmerken.
- b Welke kenmerken bij a kun je uitdrukken in een getal?
- c Vul je antwoord bij a aan zodat je ten minste drie kenmerken hebt die je niet in een getal uitdrukt.

Theorie A Enquêtevragen en statistische variabelen

In deze paragraaf richten we ons op fase 2 van de statische cyclus: Data verzamelen.

Om antwoord te vinden op de onderzoeksvraag moeten er bij een statistisch onderzoek gegevens worden verzameld.

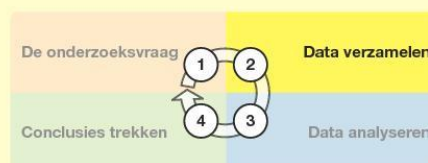
Er zijn verschillende manieren om aan gegevens te komen. Voor veel onderzoeken kan gebruik worden gemaakt van bestaand cijfermateriaal. Op internet zijn veel gegevens te vinden, bijvoorbeeld op de site van het Centraal Bureau voor de Statistiek (CBS).

Voor sommige onderzoeken zullen de juiste gegevens echter niet beschikbaar zijn. De onderzoeker zal dan bijvoorbeeld door middel van een enquête, een simulatie of een experiment de gegevens moeten verzamelen.

Een enquête is een veelgebruikte methode om gegevens te verzamelen.

Het opstellen van een goede vragenlijst is echter niet zo makkelijk.

Hieronder zie je een aantal tips van het CBS voor het maken en afnemen van een enquête. De deelnemers aan een enquête worden **respondenten** genoemd.



INFORMATIEF

Tips van het CBS voor enquêtevragen

- 1 Stel begrijpelijke vragen in korte zinnen, met duidelijke begrippen.
- 2 Stel geen suggestieve vragen zoals *‘U vindt het hier in de straat toch ook rommelig?’*.
- 3 Stel geen dubbele vragen waarbij in één vraag meerdere onderwerpen aan bod komen.
- 4 Vermijd (dubbele) ontkenningen, zoals *‘Vind je het geen probleem dat je hier niet mag kletsen?’*.
- 5 Pas op met vragen over gevoelige onderwerpen.
- 6 Pas op met vragen die een beroep doen op de herinnering zoals *‘Hoe vaak bent u de afgelopen drie jaar bij een arts of andere hulpverlener geweest?’*.
- 7 Vermijd hypothetische vragen zoals *‘Wat zou u doen als u directeur van Albert Heijn was?’*.
- 8 Geef eenduidige en complete antwoordmogelijkheden bij meerkeuzevragen.

In opgave 14 heb je een aantal kenmerken van een sporter benoemd. Deze kenmerken heten **variabelen**. Na het opstellen van de onderzoeksvraag maak je een overzicht van alle relevante statistische variabelen bij het onderzoek. Bij het onderzoek noteer je de waarde van deze variabelen voor elk element van de steekproef. Het geheel van al deze waarden zijn de data.

Binnen de statistiek onderscheiden we twee typen variabelen: kwalitatieve variabelen en kwantitatieve variabelen.

Kwalitatieve variabelen geven een kwaliteit of eigenschap weer, zoals beroep en godsdienst. Een kwalitatieve variabele is vaak niet uitgedrukt in een getal. Is dit toch het geval, dan heeft het verschil van twee opeenvolgende getallen geen eenduidige betekenis. Denk bijvoorbeeld aan de rugnummers van voetballers.

Bij **kwantitatieve variabelen** gaat het om meetbare gegevens, zoals temperatuur, snelheid en gewicht.



Bij kwantitatieve variabelen maken we onderscheid tussen discrete variabelen en continue variabelen. Een **discrete variabele** is een variabele waarbij tussenliggende waarden ontbreken, zoals bijvoorbeeld bij de variabele schoenmaat. Een **continue variabele** is een variabele die alle tussenliggende waarden kan aannemen, zoals bijvoorbeeld de variabele gewicht of snelheid.

15



a Lever commentaar op de volgende enquêtevragen.

- 1 U heeft zeker ook veel geluidsoverlast van burenen?
- 2 Vindt u niet, dat het verstandig is om geen huis te kopen als u er ook een kunt huren?
- 3 Kijkt u veel televisie?
- 4 Hoe vaak heeft u vorige week de website NU.nl bezocht?

Kies uit

I elke dag

II nooit

III 1 of 2 keer

IV 3 t/m 7 keer.

b Stel bij elk van de vier enquêtevragen van a een nieuwe vraag op waarbij je rekening houdt met alle tips van het CBS.

16



Geef bij de volgende variabelen aan of ze kwalitatief of kwantitatief zijn.

- a geslacht
- b leeftijd in jaren
- c profiel in de Tweede Fase
- d lidmaatschapsnummer

17

Geef aan of de variabele continu of discreet is.

- a** het aantal appels aan een appelboom **d** rapportcijfer
b het gewicht van de appels aan een appelboom **e** voetlengte
c de tijd die wordt besteed aan het leren voor een toets **f** schoenmaat

18

Marc en Petra hebben een discussie over de variabele leeftijd. Marc vindt leeftijd een discrete variabele, omdat leeftijd op documenten altijd in gehelen staat aangegeven. Petra brengt hier tegenin dat je leeftijd sinds je geboorte steeds toeneemt: als het moment van je geboorte bekend is, dan kun je zo nauwkeurig als je wilt bepalen hoe oud je bent. Wat vind jij? Licht je standpunt toe.

019

Het RIVM, Rijksinstituut voor Volksgezondheid en Milieu, heeft onderzoek gedaan naar eenzaamheid in Nederland. Daarbij zijn de variabelen geslacht en eenzaamheid gebruikt. Zie de tabel.

EENZAAMHEID VOLWASSENEN IN NEDERLAND

	eenzaamheid				totaal
	1	2	3	4	
man	385	206	60	12	663
vrouw	369	197	57	14	637
	754	403	117	26	1300

1 = niet eenzaam
 2 = matig eenzaam
 3 = ernstig eenzaam
 4 = zeer ernstig eenzaam

Je ziet dat bij het onderzoek 1300 volwassen Nederlanders waren betrokken. Hiervan waren 26 zeer ernstig eenzaam.

- a** Hoeveel procent van de ondervraagden was niet eenzaam?
b Hoeveel procent van de ondervraagde mannen was zeer ernstig eenzaam? En hoeveel procent van de ondervraagde vrouwen?
c Is de variabele geslacht kwalitatief of kwantitatief? En de variabele eenzaamheid?
d Bij de variabele geslacht is de volgorde man/vrouw niet van belang. Hoe zit dat bij de variabele eenzaamheid?

Theorie B Meetniveaus

In opgave 19 is het verband tussen de variabelen geslacht en eenzaamheid verwerkt in een **kruistabel**. Alhoewel de variabele eenzaamheid is aangegeven met de gradaties 1, 2, 3 en 4 is hier toch sprake van een kwalitatieve variabele. Er is echter wel een verschil met de kwalitatieve variabele geslacht. Bij het geslacht is de volgorde man/vrouw niet van belang, maar bij de eenzaamheid is er wel degelijk een volgorde. De variabele geslacht is op **nominaal niveau** gemeten en de variabele eenzaamheid op **ordinaal niveau**.

Een variabele die op een nominaal meetniveau is gemeten, is een **nominale variabele**.

Een variabele die op een ordinaal meetniveau is gemeten, is een **ordinale variabele**.

Behalve het nominale en het ordinale meetniveau zijn er nog twee meetniveaus: **interval** en **ratio**.

1 Nominaal meetniveau

Gaat het bij de meting van een variabele slechts om een onderscheid tussen verschillende categorieën, dan is er sprake van een *nominaal meetniveau*. In dat geval kunnen de waarden van de variabele aangegeven worden met *namen*, zodat ze van elkaar te onderscheiden zijn. Voorbeelden zijn de variabelen woonplaats, beroep en geslacht. Soms kent men aan de waarde van zo'n variabele een getal (code) toe, bijvoorbeeld: man 0, vrouw 1. Aan deze getallen mag je echter geen enkele betekenis toekennen wat betreft de volgorde of belangrijkheid.

2 Ordinaal meetniveau

Wordt er naast een onderscheid ook een *ordening* (volgorde) aangegeven, dan is er sprake van een *ordinaal meetniveau*.

Een voorbeeld is het toekennen van een aantal sterren aan een camping: een vijf-sterren camping heeft veel faciliteiten, een vier-sterren camping heeft er iets minder, enzovoort.

Ook het rangensysteem bij de krijgsmacht is hier een voorbeeld van. Ordinale variabelen kunnen in getallen worden uitgedrukt, maar noodzakelijk is dat niet. Denk aan de gradaties van eenzaamheid in opgave 19. Hiermee is wel een volgorde aangegeven, maar je kunt niet zeggen dat er een even groot verschil is tussen de gradaties 1 en 2 als tussen de gradaties 2 en 3.

3 Interval meetniveau

Is er naast een volgorde ook sprake van *gelijke verschillen*, dan is de meting op *intervalniveau*.

Hierbij zijn de waarden getallen waarbij het verschil tussen twee getallen een eenduidige betekenis heeft, bijvoorbeeld bij onze jaartelling. Op intervalniveau is er geen natuurlijk nulpunt.

Het jaar nul is op een gegeven moment gekozen. Je kunt dus niet zeggen dat de aarde in het jaar 2000 twee keer zo oud was als in het jaar 1000.

4 Ratio meetniveau

Bij een meting op *rationiveau* heeft een variabele alle eigenschappen van de vorige meetniveaus en is er bovendien een *natuurlijk nulpunt*.

Voorbeelden zijn de variabelen gewicht in kg en leeftijd in jaren. Bij leeftijd is de geboorte het natuurlijke nulpunt. Je kunt daarom zeggen dat een persoon van 80 jaar twee keer zo oud is als een persoon van 40 jaar.

OVERZICHT MEETNIVEAUS

nominaal	ordinaal	interval	ratio
naamgeving	naamgeving	naamgeving	naamgeving
	ordening	ordening	ordening
		gelijke verschillen	gelijke verschillen
			natuurlijk nulpunt

20
□ ◎ *

Hieronder zie je een gedeelte van het kennismakingsformulier dat een leerling moet invullen voor de mentor.

Van welk meetniveau is er bij elk van de variabelen sprake?

21
□ ◎ *

- a Bij welke meetniveaus heb je met een kwalitatieve variabele te maken? En bij welke meetniveaus met een kwantitatieve variabele?
- b Er wordt onderzoek gedaan naar de vakantiebestemming van Nederlanders in de herfstvakantie.

Bedenk in de context van dit onderzoek een variabele bij elk meetniveau.

022
□ ◎ *

Voorafgaand aan de Tweede Kamerverkiezingen in 2017 is een peiling gedaan, waarbij onderzoekers probeerden het aandeel stemmen per partij te voorspellen. Van de 1200 ondervraagden gaven 205 aan op de VVD te gaan stemmen. Bij de verkiezingen haalde de VVD uiteindelijk 21,3% van de stemmen.

Is dat meer of minder dan je op grond van de peiling zou verwachten? Licht toe.

Theorie C Populatieproportie en steekproefproportie

Bij een onderzoek onder de stemgerechtigde Nederlanders wil men de mensen uit de steekproef die lid zijn van een politieke partij enkele vragen stellen over het lidmaatschap.

Bekend is dat 2,2% van alle stemgerechtigde Nederlanders lid is van een politieke partij. We zeggen dat de **populatieproportie** gelijk is aan 0,022. De populatieproportie wordt vaak aangegeven met de letter p . De populatieproportie is dus het gedeelte van de populatie dat een bepaald kenmerk heeft.

Bij een steekproef van 1400 volwassenen blijken er 35 lid te zijn van een politieke partij. We zeggen dat de

steekproefproportie $\hat{p}$ gelijk is aan $\hat{p} = \frac{35}{1400} = 0,025$.

Bij een representatieve steekproef zal $\hat{p}$ ongeveer gelijk zijn aan p .

Soms wordt het dakje op de p bij de steekproefproportie weggelaten. Het is dan belangrijk dat je zelf goed kunt inschatten of je met p of $\hat{p}$ te maken hebt.

De populatieproportie p van een kenmerk in de populatie is

$$p = \frac{\text{aantal elementen met het kenmerk in de populatie}}{\text{totaal aantal elementen in de populatie}}.$$

De steekproefproportie $\hat{p}$ van een kenmerk in de steekproef is

$$\hat{p} = \frac{\text{aantal elementen met het kenmerk in de steekproef}}{\text{totaal aantal elementen in de steekproef}}.$$

23



De dagproductie literpakken melk in een melkfabriek is 10 000. Hiervan hebben 120 pakken te weinig inhoud. In een steekproef worden 250 pakken van de dagproductie onderzocht op de inhoud. Hiervan blijken er 4 te weinig inhoud te hebben.

Bereken de populatieproportie en de steekproefproportie bij literpakken melk met te weinig inhoud.

A24



Van de patiënten in een ziekenhuis is de populatieproportie van het kenmerk 'de patiënt heeft bloedgroep A' gelijk aan 0,418. Het blijkt namelijk dat 627 patiënten bloedgroep A hebben.

Bij een steekproef in dit ziekenhuis zijn er 27 patiënten met bloedgroep A, waarmee de steekproefproportie op 0,45 komt.

Bereken het aantal patiënten in het ziekenhuis en het aantal patiënten in de steekproef.

E25



Voor het personeel van een ziekenhuis geldt dat de populatieproportie voor het kenmerk 'personeelslid is man' gelijk is aan 0,3. Voor het kenmerk 'man heeft kinderen' geldt $p = 0,6$. Voor het kenmerk 'vrouw heeft kinderen' geldt $p = 0,5$.

Wat is de populatieproportie van het kenmerk 'personeelslid met kinderen is man'? Rond af op twee decimalen.

Terugblik

Data verzamelen

Na het formuleren van de onderzoeksvraag maak je een overzicht van alle relevante statistische variabelen bij het onderzoek. Voorbeelden zijn geslacht, leeftijd en bloedgroep. Bij het onderzoek houd je de waarde van deze variabelen bij voor elk element van de steekproef. Het geheel van al deze waarden zijn de data.

Type en meetniveau van een variabele

Bij een kwalitatieve variabele gaat het om een eigenschap of kwaliteit. Voorbeelden zijn geslacht, bloedgroep en kleur ogen.

Een kwantitatieve variabele hoort bij meetbare gegevens en wordt uitgedrukt in een getal. Voorbeelden zijn lengte, snelheid en tijd.

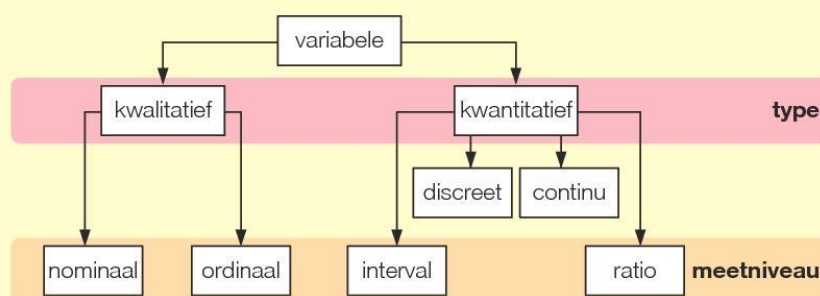
Bij variabelen onderscheiden we vier meetniveaus.

OVERZICHT MEETNIVEAUS

nominaal	ordinaal	interval	ratio
naamgeving	naamgeving	naamgeving	naamgeving
	ordening	ordening	ordening
		gelijke verschillen	gelijke verschillen
			natuurlijk nulpunt

Variabelen op nominaal en ordinaal meetniveau zijn kwalitatieve variabelen en variabelen op interval en ratio meetniveau zijn kwantitatieve variabelen.

Een kwantitatieve variabele kan discreet of continu zijn. Een discrete variabele neemt alleen 'losse' waarden aan. Bij een continue variabele kunnen alle tussenliggende waarden ook aangenomen worden.



Populatieproportie en steekproefproportie

De populatieproportie p is het deel van de populatie met het te onderzoeken kenmerk.

De steekproefproportie $\hat{p}$ is het deel van de steekproef met het te onderzoeken kenmerk.

Bij een representatieve steekproef zal $\hat{p}$ ongeveer gelijk zijn aan p .

2.3 Centrum- en spreidingsmaten

026



In de tabel zijn gegevens verwerkt over de wijze waarop de werknemers van de firma Verschut naar hun werk gaan.

- Wat is het meetniveau van de variabele 'soort vervoer'?
- Kun je bij de tabel het gemiddelde berekenen?
- Welke centrummaat kun je gebruiken om de waarnemingen te karakteriseren?

VERVOER HUIS-WERK VAN WERKNEMERS VERSCHUT

soort vervoer	bus	fiets	auto	lopen
frequentie	9	5	6	4

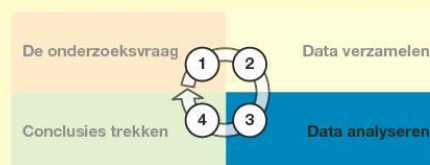
Theorie A Centrummaten

In deze paragraaf richten we ons op fase 3 van de statistische cyclus: Data analyseren.

Bij het analyseren van data kan het nuttig zijn een serie waarnemingsgetallen te typeren door één getal. Voor dat getal nemen we vaak het **gemiddelde**.

Maar het gemiddelde geeft niet altijd een juiste indruk van de waarnemingsgetallen. **Uitschieters** kunnen het gemiddelde sterk beïnvloeden. Je kunt dan beter gebruikmaken van de **mediaan** of van de **modus**.

Gemiddelde, mediaan en modus heten **centrummaten**.



Zit er een uitschieter bij de waarnemingsgetallen, dan heeft de mediaan de voorkeur boven het gemiddelde.

CENTRUMMATEN

gemiddelde	Het gemiddelde is $\frac{\text{som van de waarnemingsgetallen}}{\text{totale frequentie}}$.
mediaan	Schrijf de getallen in volgorde van grootte. De mediaan is het middelste getal bij een oneven aantal getallen. Bij een even aantal getallen is de mediaan het gemiddelde van de middelste twee getallen.
modus	De modus is het waarnemingsgetal met de grootste frequentie. Zijn er twee of meer waarnemingsgetallen met de grootste frequentie, dan is er geen modus.

oneven aantal getallen	even aantal getallen
$\underbrace{6 \ 8 \ 8 \ 10}_{\text{12}} \ \underbrace{12 \ 14 \ 15 \ 15}_{\text{12}}$ mediaan = 12	$\underbrace{3 \ 3 \ 4}_{\text{5}} \ \underbrace{5 \ 6 \ 6 \ 7 \ 8}_{\text{6}}$ mediaan = $\frac{5 + 6}{2} = 5,5$

Het hangt van de situatie af welke centrummaat het meest geschikt is. Hieronder staan enkele voor- en nadelen van de drie centrummaten.

VOORDELEN EN NADELEN VAN DE CENTRUMMATEN

	voordeel	nadeel
modus	• de enige centrummaat die bij kwalitatieve gegevens te gebruiken is	• is niet altijd aanwezig • een kleine verandering kan een geheel andere modus opleveren
mediaan	• niet gevoelig voor uitschieters	• alleen de grootte van de middelste waarneming(en) is van belang
gemiddelde	• alle gegevens worden gebruikt	• gevoelig voor uitschieters

Voorbeeld

Bij Italiaans restaurant Saponi kun je in het weekend alleen dineren na reservering. In de tabel zie je voor een weekend het aantal personen per reservering.

RESERVERINGEN SAPORI

aantal personen	1	2	3	4	5	6	10
frequentie	6	22	8	8	4	3	5

Bereken het gemiddelde, de modus en de mediaan. Rond zo nodig af op één decimaal.

Uitwerking

totale frequentie = $6 + 22 + 8 + 8 + 4 + 3 + 5 = 56$

gemiddelde =

$$\frac{6 \cdot 1 + 22 \cdot 2 + 8 \cdot 3 + \dots + 5 \cdot 10}{56} \approx 3,5$$

$$\text{mediaan} = \frac{28^{\text{e}} \text{ getal} + 29^{\text{e}} \text{ getal}}{2} = \frac{2 + 3}{2} = 2,5$$

modus = 2 Het aantal 2 heeft de grootste frequentie.

$$\begin{array}{c|c|c|c|c|c|c|c} 1 & 2 & 3 & 4 & 5 & 6 & 10 \\ \hline 6 & 22 & 8 & 8 & 4 & 3 & 5 \\ \hline \end{array}$$

$$\underbrace{6+22}_{28} + \underbrace{8+8+4+3+5}_{28} = 56$$

getallen getallen

$$\left. \begin{array}{l} 28^{\text{e}} \text{ getal} = 2 \\ 29^{\text{e}} \text{ getal} = 3 \end{array} \right\} \text{ mediaan} = 2,5$$

27
□ ⊙ *

In klas 4VA is gedurende enige tijd dagelijks bijgehouden hoeveel leerlingen die dag te laat in de les kwamen.

TELAATKOMERS IN KLAS 4VA

aantal	0	1	2	3	4	5	6
frequentie	3	3	7	0	3	2	2

a Hoeveel dagen duurde het onderzoek?

b Bereken het gemiddelde, de mediaan en de modus.



28

Ook in klas 4VB is nauwgezet het aantal telaatkomers geregistreerd.



TELAATKOMERS IN 4VB

aantal	0	1	2	3	4	5	32
frequentie	1	5	5	3	3	2	1

- a** Bereken het gemiddelde.
- b** Wanneer de afdelingscoördinator het gemiddelde van klas 4VB vergelijkt met dat van klas 4VA uit opgave 27 is zij erg ontstemd. Ze vindt het verschil onaanvaardbaar hoog. Door het vergelijken van een andere centrummaat weet klas 4VB de afdelingscoördinator milder te stemmen.
Hoe doet de klas dat?

29



In de tabel zie je de resultaten van Jim en Nils op hun rapport. Jim heeft Duits als extra vak, waardoor hij één rapportcijfer meer heeft dan Nils. Voor de vakken CKV en LO kregen ze geen rapportcijfer, maar een woordbeoordeling.

- a** Bereken de mediaan van de rapportcijfers van Jim.
- b** Bereken de mediaan van de rapportcijfers van Nils.
- c** Nils wil zorgen dat de mediaan van zijn rapportcijfers hoger wordt dan een 7,0.
Op welk vak kan hij zich dan het beste richten? Licht toe met een berekening.

vak	beoordeling	
	Jim	Nils
NE	5,9	6,6
EN	8,4	7,4
DU	8,0	-
WIA	7,2	7,3
SK	6,3	5,9
BIO	6,8	6,3
NLT	7,0	7,1
AK	6,6	5,7
MAA	7,0	7,0
CKV	V	V
LO	V	V

30



Welke centrummaat is het meest geschikt om de volgende waarnemingen te karakteriseren?

- a** Op een verjaardagsfeestje hebben de aanwezigen de volgende leeftijden: 17, 20, 22, 17, 19, 21 en 68 jaar.
- b** Van tien leerlingen is de haarkleur: blond, rood, bruin, bruin, zwart, bruin, bruin, blond, bruin en blond.
- c** Een verkoper verkoopt op de zes werkdagen van een week de volgende aantallen fietsen: 8, 7, 4, 5, 8 en 6.
- d** Bij een natuurkundeproef meten tien leerlingen het volume van dezelfde glazen bol. De metingen leveren op: 48, 48, 48, 48, 53, 48, 48, 48, 48 en 48 cm³.



A31
□ ⊗ *

Om een vwo-diploma te behalen moet je aan allerlei eisen voldoen. Eén van die eisen is dat je gemiddeld een 5,50 of hoger haalt voor de vakken waarin je centraal examen doet.

Daan en Tim hebben examen gedaan. Daan heeft voor de zeven vakken waarin hij centraal examen deed, gemiddeld een 5,06 gehaald. Tim, die in acht vakken centraal examen deed, haalde gemiddeld een 5,08. Verder voldoen ze aan alle eisen die gesteld worden om een vwo-diploma te behalen. Daan en Tim krijgen de kans om een willekeurig vak te herkansen, om zodoende op een gemiddelde van 5,50 of hoger uit te komen. Bereken wie van de twee de grootste cijfermatige verbetering moet behalen om alsnog voor het vwo-examen te slagen.

E32
*

Bij een toets in klas V4a zijn er 50 punten te behalen. Na afloop van de toets bleek dat de gemiddelde score 1,2 punten hoger zou zijn geweest als iedere jongen van de klas 3 punten meer had gehaald. Hoeveel procent van de klas is meisje?

O33
□ ⊗ *

- In de tabel hiernaast zie je hoeveel nachten 60 gasten op camping De Flierweide verbleven.
- a** Wat is het meetniveau van de variabele 'aantal nachten'?
 - b** Hoeveel nachten verbleef de 25% langst verblijvende gasten?
 - c** Hoeveel nachten verbleven de gasten die wat betreft het aantal overnachtingen de middelste 50% vormen?

OVERNACHTINGEN CAMPING

aantal nachten	frequentie
1	3
2	1
3	6
4	5
5	9
6	13
7	8
10	9
12	4
14	2

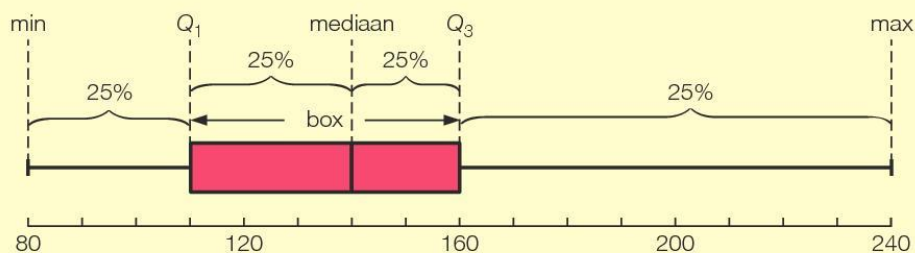
Theorie B De boxplot

Een **boxplot** is een diagram dat de verdeling van data in beeld brengt. Alleen bij een kwantitatieve variabele kun je een boxplot maken.

De mediaan verdeelt een serie waarnemingsgetallen in twee groepen, die beide bestaan uit 50% van de waarnemingsgetallen. Van beide groepen kun je weer de mediaan bepalen. Er ontstaat dan een verdeling in vier groepen, die elk 25% van de waarnemingsgetallen bevatten.

De mediaan van de eerste helft heet het **eerste kwartiel**. Notatie Q_1 .

De mediaan van de tweede helft heet het **derde kwartiel**. Notatie Q_3 .



figuur 2.6

In de boxplot hierboven lees je af dat

- de 25% kleinste waarnemingsgetallen ligt tussen 80 en 110
- de volgende 25% van de waarnemingsgetallen ligt tussen 110 en 140
- de volgende 25% van de waarnemingsgetallen ligt tussen 140 en 160
- de 25% grootste waarnemingsgetallen ligt tussen 160 en 240.

Elk deel van de boxplot bevat 25% van de waarnemingsgetallen.

Het verschil tussen het grootste en kleinste waarnemingsgetal heet de **spreidingsbreedte**.

In de figuur is de spreidingsbreedte $240 - 80 = 160$.

De breedte van de **box** in de boxplot heet de **interkwartielafstand** (ook wel kwartielafstand).

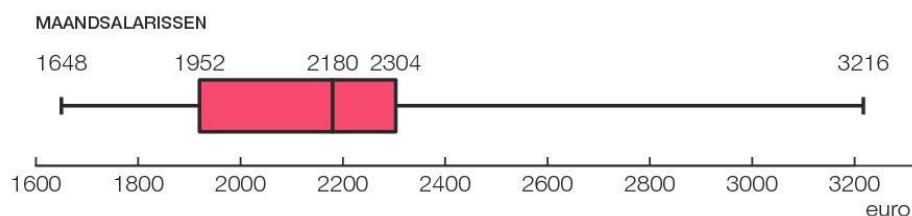
In de figuur is de interkwartielafstand $160 - 110 = 50$.

spreidingsbreedte = maximum – minimum

interkwartielafstand = $Q_3 - Q_1$

34
📊

Van de 250 werknemers van een ICT-bedrijf is het maandsalaris in een boxplot verwerkt. Zie de figuur hieronder.

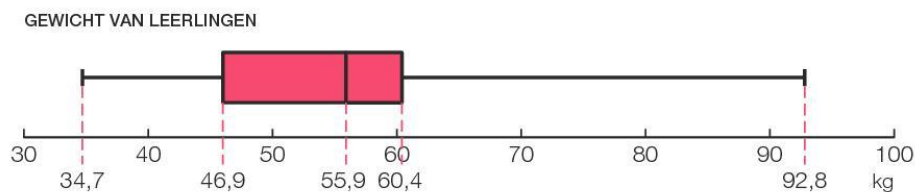


figuur 2.7

- Bereken de spreidingsbreedte en de interkwartielafstand.
- Hoeveel werknemers hebben een maandsalaris tussen 1952 en 2304 euro?
- Wat weet je van de maandsalarissen van de 25% werknemers met de hoogste maandsalarissen?

35
□ ⊗ *

Van een grote scholengemeenschap zijn de gewichten van alle 2184 leerlingen genoteerd. Er is bij deze data een boxplot gemaakt. Zie figuur 2.8.



figuur 2.8

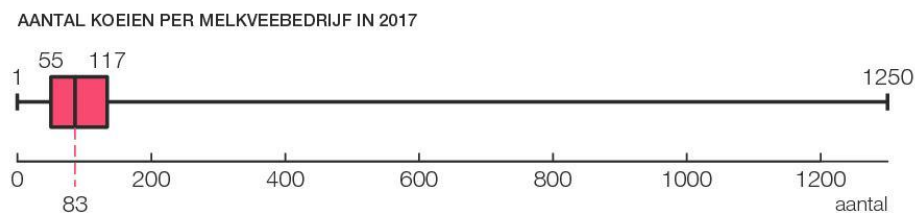
Een leerling trekt uit bovenstaande gegevens de volgende conclusies.

Geef bij elke conclusie aan of deze juist is of onjuist. Licht toe.

- Er zijn meer leerlingen die een gewicht hebben tussen 60,4 en 92,8 kg dan tussen 55,9 en 60,4 kg.
- Er zijn meer dan 550 leerlingen die lichter zijn dan 46,9 kg.
- Er vallen minder leerlingen binnen dan buiten de box.
- Meer dan driekwart van de leerlingen is lichter dan het gemiddelde van het kleinste en grootste gewicht.

A36
□ ⊗ *

Er wordt al jaren onderzoek gedaan naar het aantal melkkoeien per melkveebedrijf in Nederland.



figuur 2.9

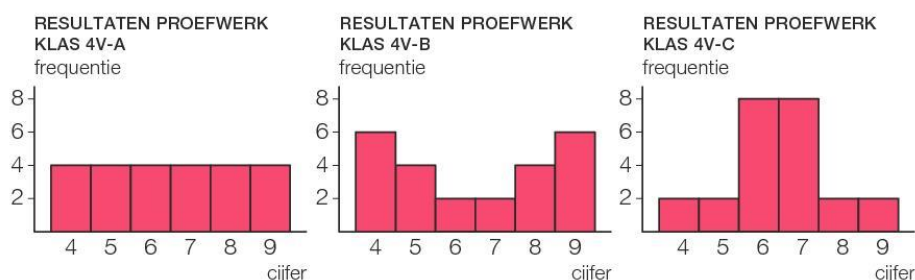
Zie de boxplot. In 2017 waren er 18056 melkveebedrijven.

- Hoeveel bedrijven hebben meer dan 83 koeien?
- Is het gemiddelde hoger of lager dan de mediaan? Licht toe.
- Verbeter de volgende bewering zodat deze correct wordt.
Er zijn meer bedrijven met meer dan 117 koeien dan bedrijven met minder dan 55 koeien.
- Onderzoek of de volgende bewering kan kloppen.
De bedrijven met meer dan 117 koeien hebben samen meer koeien dan de bedrijven met minder dan 117 koeien.
- Geef een schatting van het aantal bedrijven met minder dan 100 koeien.



Drie klassen maken hetzelfde proefwerk. Van elke klas is het gemiddelde cijfer 6,5 en de spreidingsbreedte 5. Toch is er verschil in spreiding, zie figuur 2.10.

Omschrijf dit verschil.



figuur 2.10

Theorie C De standaardafwijking

Twee groepen met gelijke mediaan of gelijk gemiddelde kunnen nog steeds sterk verschillen. Om nog meer informatie te geven over een groep kun je gebruikmaken van een **spreidingsmaat**. Een spreidingsmaat drukt de mate van spreiding van de gegevens uit in een getal.

De eenvoudigste spreidingsmaat is de spreidingsbreedte, maar deze is zeer gevoelig voor uitschieters.

De interkwartielafstand heeft dit nadeel niet en is een voorbeeld van een spreidingsmaat die vaak in combinatie met de mediaan wordt gebruikt.

De meest gebruikte spreidingsmaat is de **standaardafwijking**, ook wel de **standaarddeviatie** genoemd.

De standaardafwijking wordt vaak in combinatie met het gemiddelde gebruikt. De gebruikelijke notatie voor de standaardafwijking is σ , dat is de Griekse letter sigma. Net als het gemiddelde maakt de standaardafwijking gebruik van alle waarnemingsgetallen.

De standaardafwijking geeft een indruk van de gemiddelde afwijking van de waarnemingsgetallen tot het gemiddelde.

De standaardafwijking met de hand berekenen is een hele klus en daarom wordt dat bijna altijd elektronisch gedaan. Om de standaardafwijking met de hand te berekenen, ga je als volgt te werk:

- 1 Bereken het gemiddelde van de waarnemingsgetallen.
- 2 Bereken voor elk waarnemingsgetal het verschil met het gemiddelde, dus waarnemingsgetal – gemiddelde.
- 3 Tel de kwadraten van al deze verschillen bij elkaar op.
- 4 Deel de uitkomst door het aantal waarnemingsgetallen.
- 5 Neem de wortel van deze uitkomst. Dit getal is de standaardafwijking van de waarnemingsgetallen.

38



Coen heeft de volgende getallen in zijn schrift staan.

rij A:	0, 2, 3, 4, 6
rij B:	10, 12, 13, 14, 16
rij C:	3, 3, 3, 3, 3, 3

figuur 2.11

- Bereken de standaardafwijking van rij A.
- Hoe komt het dat de standaardafwijkingen van de rijen A en B gelijk zijn?
- Geef de standaardafwijking van rij C.

39



Meneer Boers heeft in zijn klas een toets afgenomen. De cijfers variëren van 2,8 tot en met 9,4. Het gemiddelde is 6,2.

- Hij berekent $\sigma = 5,3$.
Licht toe waarom dit niet kan kloppen.

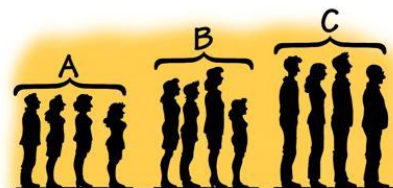
Na het bespreken van de toets blijkt dat meneer Boers een fout heeft gemaakt bij het tellen van de punten. Het cijfer 2,8 moet een 3,9 zijn.

- Is de standaardafwijking hierdoor groter of kleiner geworden? Licht toe.

A40

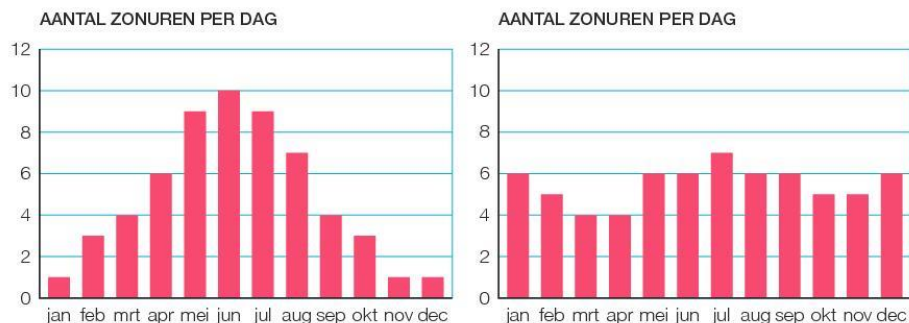


- Bij welke van de groepen hiernaast is de standaardafwijking van de lengte het grootst? Licht toe.



figuur 2.12

- In onderstaande figuren kun je voor iedere maand het gemiddeld aantal zonuren per dag in de steden Helsinki en Quito aflezen. De standaardafwijking bij Helsinki is groter dan bij Quito. Welke van de twee figuren hoort bij Quito? Licht toe.



a

b

figuur 2.13

Terugblik

Centrummaten

Om de verzamelde data te karakteriseren met één getal kun je een centrummaat gebruiken.

CENTRUMMATEN

gemiddelde	mediaan	modus
$\frac{\text{som van de getallen}}{\text{totale frequentie}}$	middelste getal (of gemiddelde van de middelste twee) in een rangschikking naar grootte	waarnemingsgetal met de grootste frequentie
alle gegevens worden gebruikt	alleen de grootte van de middelste waarneming(en) is van belang	ook bruikbaar bij kwalitatieve gegevens
gevoelig voor uitschieters	niet gevoelig voor uitschieters	gevoelig voor kleine veranderingen

Kwartielen

De mediaan verdeelt een serie waarnemingsgetallen in twee groepen, die beide bestaan uit 50% van de waarnemingsgetallen. Van elk van deze even grote groepen kun je weer de mediaan bepalen. Er ontstaat dan een verdeling in vier groepen, die elk 25% van de waarnemingsgetallen bevatten.

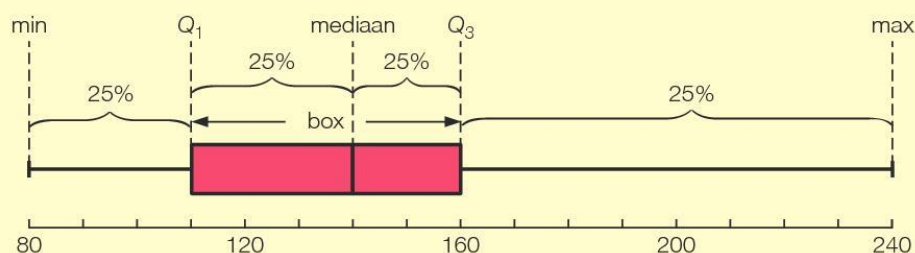
De mediaan van de eerste helft heet het eerste kwartiel. Notatie Q_1 .

De mediaan van de tweede helft heet het derde kwartiel. Notatie Q_3 .

Boxplot

Een boxplot is een diagram dat de spreiding van data in beeld brengt.

Alleen bij een kwantitatieve variabele kun je een boxplot maken.



Spreidingsmaten

Je kunt de volgende spreidingsmaten gebruiken:

- spreidingsbreedte = grootste waarnemingsgetal – kleinste waarnemingsgetal
- interkwartielafstand = derde kwartiel – eerste kwartiel
- standaardafwijking σ

De standaardafwijking gebruikt men vaak in combinatie met het gemiddelde, bij de mediaan is de logische keuze de interkwartielafstand.

2.4 Klassenindeling en polygonen

041
□ ⊗ *

Anneke heeft een tijdje bijgehouden hoeveel euro zij uitgeeft in de supermarkt. De bedragen, afgerond op gehele euro's, staan hieronder.

55	28	36	68	21	17	65
32	20	26	15	58	11	43
57	69	33	27	47	38	56
29	25	14	18	16	51	44

Waarom is het niet zinvol hierbij een frequentietabel te maken?



Theorie A Klassenindeling

Ook in deze paragraaf richten we ons op fase 3 van de statistische cyclus: Data analyseren.

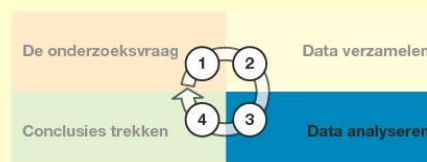
In opgave 41 komt elk waarnemingsgetal maar één keer voor. Het heeft dan geen zin een frequentietabel te maken, want alleen de frequentie 1 komt voor. Om een beter inzicht te krijgen in de bedragen kun je de bedragen indelen in groepen. Deze groepen heten **klassen**. Omdat het kleinste bedrag 11 euro is en het grootste 69 euro, is een verstandige klassenindeling $[10, 20)$, $[20, 30)$, $[30, 40)$, $[40, 50)$, $[50, 60)$ en $[60, 70)$.

Bij de klasse $[10, 20)$ geef je met $[$ aan dat 10 wel tot de klasse behoort en met $)$ dat 20 niet tot de klasse behoort.

Bij deze klasse heten 10 en 20 de **klassengrenzen**.

De **klassenbreedte** is $20 - 10 = 10$.

Het **klassenmidden** is $\frac{10 + 20}{2} = 15$.



De klasse $[10, 20)$ loopt van 10 tot 20.
Bij $\langle 10, 20]$ hoort 10 niet bij de klasse en 20 wel.

Aan een frequentietabel met een klassenindeling is niet meer te zien van welke waarnemingsgetallen is uitgegaan.

Om een benadering te krijgen van het gemiddelde kun je gebruikmaken van de klassenmiddens.

In plaats van de modus spreken we bij een klassenindeling van de **modale klasse**, dat is de klasse met de grootste frequentie.

Ook de mediaan is bij een klassenindeling niet meer te berekenen. Je kunt wel berekenen in welke klasse de mediaan ligt.

Voorbeeld

Van een aantal leerlingen is het gewicht bepaald.

Zie de frequentietabel hiernaast.

- a Bereken het gemiddelde gewicht met behulp van klassenmiddens.
- b In welke klasse ligt de mediaan?
- c Geef de modale klasse.

GEWICHT LEERLINGEN IN KG

gewicht	frequentie
[50, 55)	4
[55, 60)	7
[60, 65)	12
[65, 70)	14
[70, 75)	5
[75, 80)	8

Uitwerking

- a totale frequentie = $4 + 7 + 12 + 14 + 5 + 8 = 50$

$$\text{gemiddelde} = \frac{4 \cdot 52,5 + 7 \cdot 57,5 + \dots + 8 \cdot 77,5}{50} = 65,8 \text{ kg}$$

- b De mediaan is het gemiddelde van het 25^e en 26^e getal, dit ligt in de klasse [65, 70). ←.....
- c De modale klasse is [65, 70).

In de eerste drie klassen zitten $4 + 7 + 12 = 23$ getallen.

42



Van een aantal levensmiddelen in een supermarkt is de prijs genoteerd. Zie de frequentietabel hiernaast.

- a Heb je hier te maken met een continue of een discrete variabele? Licht toe.
- b Een zak snoep kost € 3,00. In welke klasse valt dit?
- c Bereken de gemiddelde prijs met behulp van klassenmiddens.
- d In welke klasse ligt de mediaan?
- e Geef de modale klasse.

PRIJS LEVENSMIDDELEN

prijs in centen	frequentie
<0, 100]	26
<100, 200]	17
<200, 300]	21
<300, 400]	8
<400, 500]	13
<500, 600]	2

A43



Bij een onderzoek in een winkelcentrum is van elke winkel het aantal werknemers en het aantal m² vloeroppervlakte vastgesteld.

winkel	1	2	3	4	5	6	7	8	9	10
aantal werknemers	6	8	18	25	7	4	22	16	12	8
vloeroppervlakte in m ²	125	85	240	325	412	72	450	315	65	82

winkel	11	12	13	14	15	16	17	18	19	20
aantal werknemers	18	23	14	9	4	8	8	28	9	19
vloeroppervlakte in m ²	180	521	228	172	88	56	185	270	120	438

- a Maak een frequentietabel bij de vloeroppervlakte in m². Neem de klassen [50, 150), [150, 200), ...
- b Wat is de modale klasse bij de vloeroppervlakte?
- c Met behulp van de gegeven tabel kun je berekenen dat de gemiddelde vloeroppervlakte per werknemer ongeveer 16,7 m² is. Bereken de gemiddelde vloeroppervlakte per werknemer door de klassenmiddens van de frequentietabel bij vraag a te gebruiken. Verklaar het verschil tussen de twee gemiddelden.

A44



Aan de docenten van het Noorderveld College is gevraagd hoe oud ze zijn. Zie de frequentietabel hiernaast.

- Mevrouw Van Santen is 35 jaar.
In welke leeftijdsklasse valt ze?
- Een docent Frans zegt dat de gemiddelde leeftijd van de docenten 40 jaar is.
Onderzoek of dit kan kloppen.
- Kan de mediaan 47 jaar zijn? Licht toe.
- Conciërge Erwin beweert dat de modus van de leeftijden 36 is.
Waarom kan dit niet kloppen?

DOCENTEN N.C.

leeftijd	frequentie
[25, 35>	18
[35, 45>	3
[45, 55>	40
[55, 65>	9
[65, 75>	2

O45



In klas 4VC is aan elke leerling gevraagd hoeveel broers en zussen hij of zij heeft. Zie de frequentietabel.

- Is de variabele discreet of continu? Wat is het meetniveau?
- Geef de praktische betekenis van het getal 6 in de tabel.
- Hoeveel leerlingen zitten er in klas 4VC?

AANTAL BROERS EN ZUSSEN

aantal	frequentie
0	5
1	12
2	6
3	2
4	2
5	1

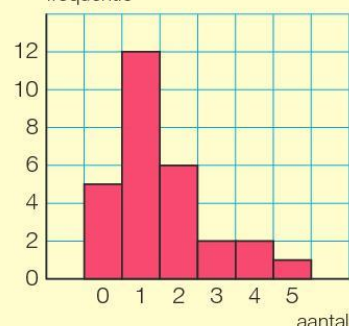
Theorie B Histogram en frequentiepolygoon

In figuur 2.14 is bij de **frequentieverdeling** van opgave 45 het **histogram** getekend. Bij een histogram staat op de horizontale as een kwantitatieve variabele. Soms, zoals bij dit histogram, staan de waarnemingsgetallen midden onder de staven. Heb je te maken met een klassenindeling, dan staan de klassengrenzen bij de grenzen van de staven. De staven in een histogram staan tegen elkaar aan.

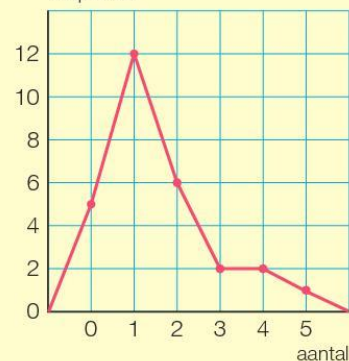
Een histogram is een diagram bij een frequentieverdeling met een kwantitatieve variabele op de horizontale as. De staven staan tegen elkaar aan.

Bij een kwalitatieve variabele kun je geen histogram maken. Je kunt wel een staafdiagram maken, de staven staan dan los van elkaar.

In figuur 2.15 is bij de frequentieverdeling van opgave 45 een lijndiagram getekend. Omdat hier geen klassenindeling is gebruikt, is boven elk waarnemingsgetal een stip gezet. De hoogte van de stip komt overeen met de frequentie van het waarnemingsgetal. Bij een klassenindeling zou die stip boven het klassenmidden geplaatst zijn. Door de opeenvolgende stippen te verbinden met lijnstukken krijg je de **frequentiepolygoon**. Je ziet dat de polygoon op de horizontale as begint en eindigt.

BROERS EN ZUSSEN 4VC
frequentie

figuur 2.14

BROERS EN ZUSSEN 4VC
frequentie

figuur 2.15

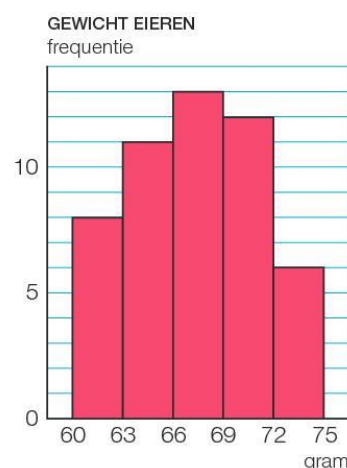
Een frequentiepolygoon is een lijndiagram waarin de frequenties zijn uitgezet tegen de waarnemingsgetallen. Het begin- en eindpunt liggen op de horizontale as.

In figuur 2.15 is de frequentiepolygoon getekend bij absolute frequenties. Je kunt ook een frequentiepolygoon tekenen bij relatieve frequenties. Je krijgt dan de **relatieve frequentiepolygoon**.

46 Van 50 eieren is het gewicht vastgesteld. De resultaten zijn verwerkt in de figuur hiernaast met als eerste klasse $[60, 63)$.



- Met welke variabele heb je hier te maken? Is deze variabele kwalitatief of kwantitatief?
- Geef de modale klasse.
- Matthijs beweert dat het totale gewicht van de eieren tussen 3291 en 3441 gram ligt. Toon aan dat hij gelijk heeft.
- Onderzoek of het mogelijk is dat het gemiddelde gewicht van de eieren buiten de klasse $[66, 69)$ ligt.



figuur 2.16

47 Boer Jopma heeft bij een partij eieren de tabel hiernaast gemaakt.



- Is de variabele klasse kwalitatief of kwantitatief? Licht toe.
- Jopma wil weten wat het gemiddelde gewicht van deze partij eieren ongeveer is. Kan hij dit berekenen met de gegevens in de tabel? Licht toe.
- Jopma wil de gegevens uit de tabel verwerken in een staafdiagram. Waarom kan hij geen histogram gebruiken?

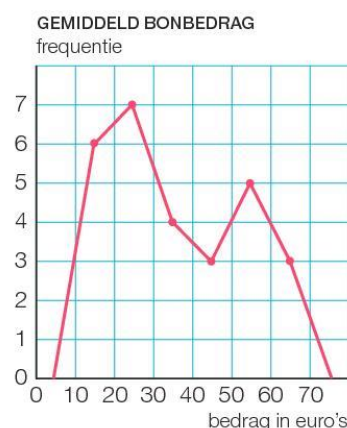
EIEREN BOER JOPMA

klasse	gewicht in gram	aantal
S	$\langle \leftarrow, 53 \rangle$	441
M	$[53, 63)$	541
L	$[63, 73)$	533
XL	$[73, \rightarrow)$	216

48 Voor een aantal winkels is de gemiddelde besteding per klant (gemiddeld bonbedrag) verwerkt in een klassenindeling met als eerste klasse $[10, 20)$. In de figuur hiernaast zie je de bijbehorende frequentiepolygoon.



- Met welke variabele heb je hier te maken? Is deze variabele discreet of continu? Licht toe.
- Wat is het laagst mogelijke bonbedrag?
- Van hoeveel winkels is het gemiddeld bonbedrag in de figuur verwerkt?
- In welke klasse ligt de mediaan?
- Bereken de relatieve frequentie van de klasse $[50, 60)$.

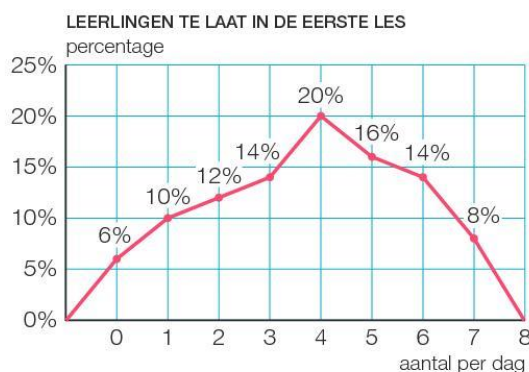


figuur 2.17

A49



De verzuimbeheerder op het Dominicus College heeft een aantal dagen bijgehouden hoeveel leerlingen te laat in hun eerste les kwamen. Tot haar schrik kwam het tijdens de onderzoeksperiode 33 dagen voor dat er meer dan vijf leerlingen te laat in hun eerste les waren. Alle data van haar onderzoek staat verwerkt in de relatieve frequentiepolygoon hiernaast.



figuur 2.18

- Is het mogelijk om bij deze gegevens een histogram te maken?
Licht toe.
- Bereken het gemiddeld aantal leerlingen dat te laat kwam in de eerste les.
- Hoeveel dagen duurde het onderzoek?
- Bereken hoe vaak de verzuimbeheerder genoteerd heeft dat een leerling te laat in de eerste les kwam.

O50



In de tabel hiernaast vind je een klassenindeling bij de lengte van mannelijke sporters op de Olympische Spelen van 2012.

LENGTE MANNELIJKE OLYMPISCHE SPORTERS

- Diego Castillo is een zwemmer met een lengte van 175 cm.
Bereken hoeveel procent van de sporters in de tabel kleiner is dan Diego.
- Camille Lacourt is een zwemmer met een lengte van 2 meter die deelnam aan deze Olympische Spelen.
Uit de tabel volgt dat Camille bij de 5% langste sporters hoort.
Laat zien dat dit klopt.

lengte	rel. freq.
$\langle \leftarrow, 155 \rangle$	0,3%
$[155, 160 \rangle$	0,7%
$[160, 165 \rangle$	2,4%
$[165, 170 \rangle$	5,9%
$[170, 175 \rangle$	12,2%
$[175, 180 \rangle$	16,3%
$[180, 185 \rangle$	20,4%
$[185, 190 \rangle$	17,2%
$[190, 195 \rangle$	13,0%
$[195, 200 \rangle$	6,9%
$[200, 205 \rangle$	3,0%
$[205, 210 \rangle$	1,1%
$[210, 215 \rangle$	0,4%
$[215, \rightarrow \rangle$	0,2%

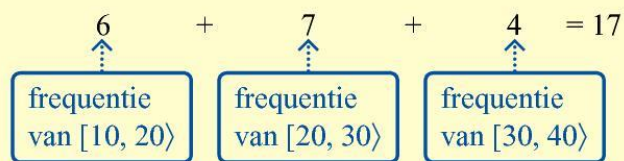
Voor het onderling vergelijken van resultaten maakt men soms gebruik van het begrip *percentielscore*. Deze score geeft in een getal weer hoeveel procent van de resultaten dezelfde waarde of lager heeft.

- Bij de lengte van een mannelijke sporter uit de tabel hoort een percentielscore van 65.
Wat weet je over zijn lengte?
- Welke percentielscore hoort bij de mediaan?

Theorie C Cumulatieve frequentiepolygoon

Tel je bij de frequentie van een waarnemingsgetal ook alle voorgaande frequenties mee, dan spreken we van de **cumulatieve frequentie**. Wanneer je met cumulatieve frequenties in procenten te maken hebt, noem je het resultaat de **relatieve cumulatieve frequentie**.

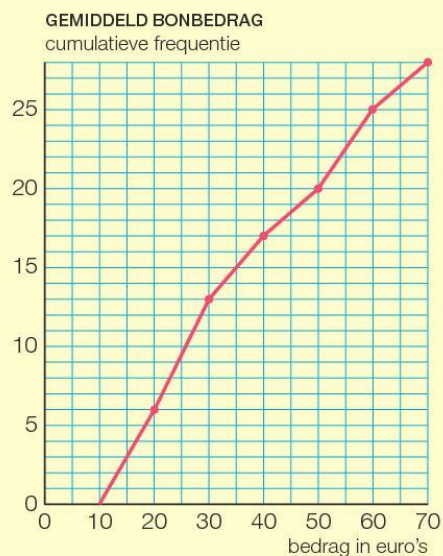
In de tabel hiernaast bereken je de cumulatieve frequentie van de klasse $[30, 40)$ door de frequentie van deze klasse én alle voorgaande klassen op te tellen. De cumulatieve frequentie van de klasse $[30, 40)$ is



Er zijn dus 17 winkels waarbij het gemiddeld bonbedrag minder dan 40 euro is.

In figuur 2.19 is de **cumulatieve frequentiepolygoon** getekend. De punten zijn uitgezet boven de *rechtergrens* van elke klasse.

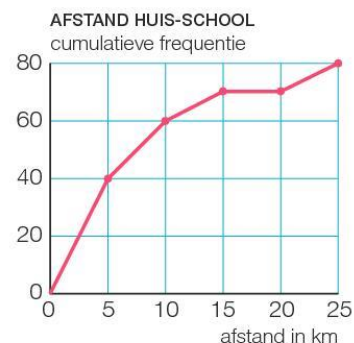
GEMIDDELD BONBEDRAG	
bedrag in euro's	frequentie
$[10, 20)$	6
$[20, 30)$	7
$[30, 40)$	4
$[40, 50)$	3
$[50, 60)$	5
$[60, 70)$	3



figuur 2.19

R51 Waarom worden (relatieve) cumulatieve frequenties in een polygoon uitgezet boven de rechterklassengrenzen?

- 52** Van een groep studenten is de afstand huis-school in km genoteerd. Zie de cumulatieve frequentiepolygoon in figuur 2.20, waarbij de eerste klasse $[0, 5)$ is.
- Hoeveel studenten zijn bij het onderzoek betrokken?
 - Hoeveel studenten wonen minder dan 10 km van school?
 - Hoeveel studenten wonen minstens 5 km van school?
 - Hoeveel studenten wonen minstens 10 km, maar minder dan 20 km van school?
 - Teken een histogram bij figuur 2.20.



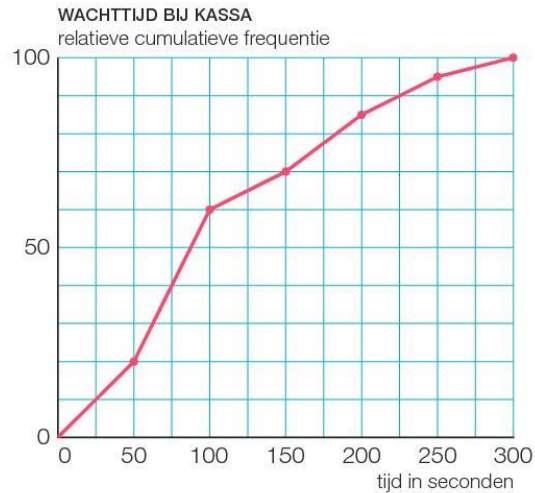
figuur 2.20

53



In een supermarkt is bijgehouden hoeveel seconden een klant in de rij bij de kassa staat tot hij aan de beurt is. De wachttijden van 300 klanten zijn verwerkt in de relatieve cumulatieve frequentiepolygoon in figuur 2.21, waarbij de eerste klasse $[0, 50)$ is.

- Hoeveel procent van de klanten stond minder dan 100 seconden in de rij? Hoeveel klanten zijn dat?
- Hoeveel klanten stonden minder dan 200 seconden in de rij?
- Hoeveel klanten moesten minstens 2,5 minuut wachten?
- Hoeveel klanten moesten minstens 100, maar minder dan 200 seconden wachten?



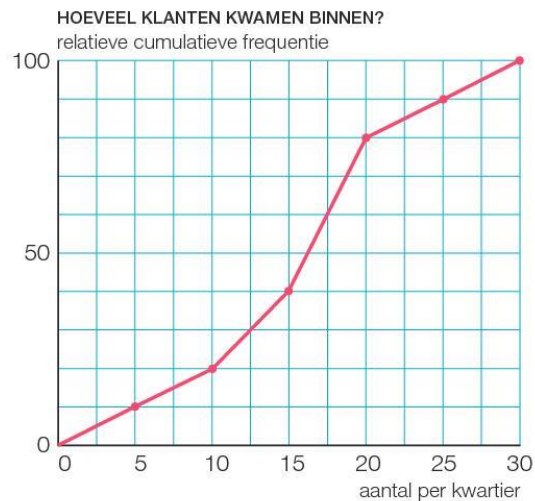
figuur 2.21

A54



In een winkel is op een dag van 10:00 tot 20:00 uur genoteerd hoeveel klanten er per kwartier de winkel binnenkwamen. Zie figuur 2.22, waarbij de eerste klasse $[0, 5)$ is.

- Hoeveel kwartieren kwamen minder dan 20 klanten binnen?
- Hoeveel kwartieren kwamen minstens 15 klanten binnen?
- Hoeveel kwartieren kwamen minstens 10, maar minder dan 25 klanten de winkel binnen?
- Maak bij figuur 2.22 een frequentietabel met absolute frequenties.



figuur 2.22

Terugblik

Klassenindeling

Zijn er bij een statistisch onderzoek veel verschillende waarnemingsgetallen, dan maak je een indeling in klassen.

Geef elke klasse dezelfde breedte.

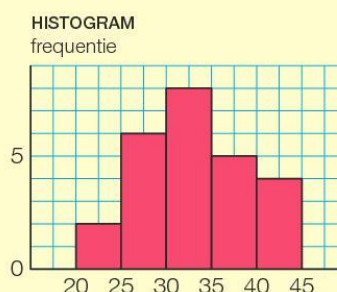
Bij de klasse $[20, 25)$ doet de linkergrens 20 mee, maar de rechtergrens 25 niet. De klassenbreedte van deze klasse is

$$25 - 20 = 5. \text{ Het klassenmidden is } \frac{20 + 25}{2} = 22,5.$$

Om bij een klassenindeling een schatting van het gemiddelde te verkrijgen, kun je gebruikmaken van de klassenmiddens.

Bij een klassenindeling kun je een histogram tekenen. Zo krijg je bij de verdeling hieronder het histogram ernaast.

klasse	frequentie
$[20, 25)$	2
$[25, 30)$	6
$[30, 35)$	8
$[35, 40)$	5
$[40, 45)$	4

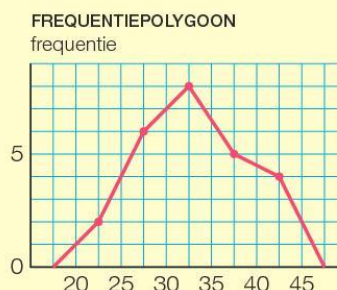


Frequentiepolygoon

Bij een frequentieverdeling van een kwantitatieve variabele is een frequentiepolygoon te maken.

Een frequentiepolygoon is een lijndiagram waarin boven elk waarnemingsgetal de frequentie is uitgezet. Begin- en eindpunt liggen op de horizontale as.

Bij een relatieve frequentiepolygoon zet je de relatieve frequenties uit tegen de waarnemingsgetallen.



Cumulatieve frequenties

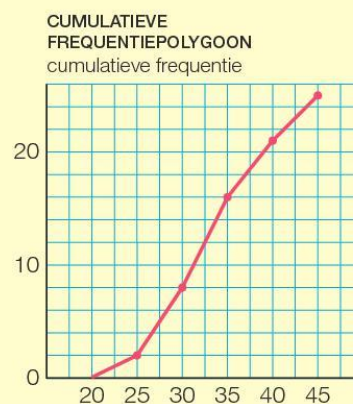
De cumulatieve frequentie van de klasse $[30, 35)$ krijg je door de frequentie van deze klasse en alle

vooraangaande klassen op te tellen. De cumulatieve frequentie van de klasse $[30, 35)$ is dus $2 + 6 + 8 = 16$.

Er zijn dus 16 waarnemingsgetallen kleiner dan 35.

De cumulatieve frequenties bij de tabel hierboven zijn achtereenvolgens 2, 8, 16, 21 en 25.

In de figuur hiernaast zie je de cumulatieve frequentiepolygoon.



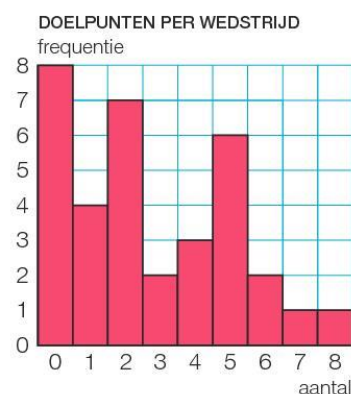
2.5 Soorten verdelingen

O55
□ ⊗ *

Bij een schoolvoetbaltoernooi is per wedstrijd het aantal doelpunten genoteerd.

Zie het histogram hiernaast.

- Wat is de betekenis van de staaf bij het aantal 2?
- Hoeveel wedstrijden zijn er gespeeld?
- Hoeveel doelpunten zijn er in totaal gescoord?
- Wat is het maximaal aantal wedstrijden dat geëindigd is in 2-2?
- Onderzoek of het gemiddeld aantal doelpunten per wedstrijd groter of kleiner is dan de mediaan.

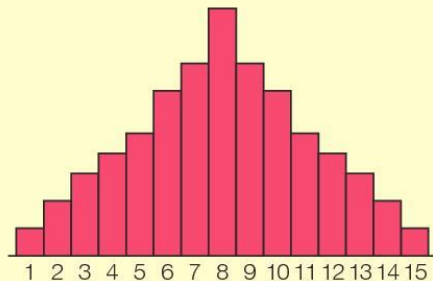
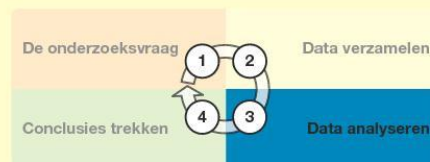


figuur 2.23

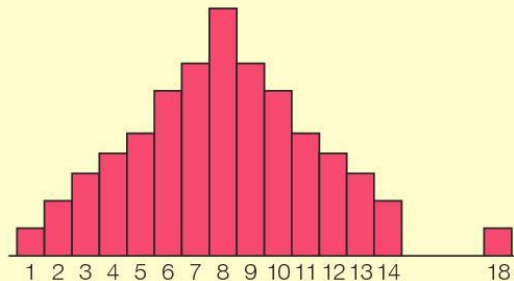
Theorie A Verdelingen en histogrammen

Ook in deze paragraaf richten we ons op fase 3 van de statistische cyclus: Data analyseren.

Het histogram in figuur 2.24 hoort bij een **symmetrische verdeling**. Het gemiddelde, de mediaan en de modus vallen samen.



figuur 2.24



figuur 2.25

In het histogram van figuur 2.25 zie je aan de rechterkant een uitschieter. Merk op dat het staafje dat in figuur 2.24 bij 15 staat, in figuur 2.25 bij 18 staat. Het gemiddelde zal bij figuur 2.25 groter zijn dan bij figuur 2.24, maar de modus is in beide figuren gelijk.

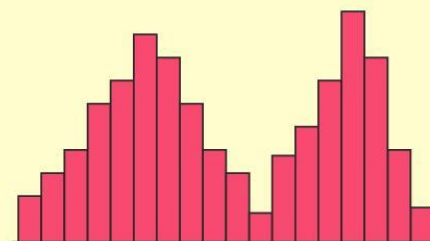
Omdat in beide figuren evenveel waarnemingsgetallen staan en de volgorde van de waarnemingsgetallen van klein naar groot hetzelfde is, hebben de verdelingen dezelfde mediaan.

In figuur 2.25 is dus het gemiddelde groter dan de mediaan.

In figuur 2.25 is door de uitschieter de spreiding groter dan in figuur 2.24. Daarom zal in figuur 2.25 de standaardafwijking groter zijn dan in figuur 2.24.

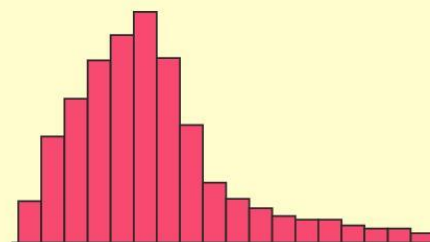
In de figuren hiernaast zijn histogrammen getekend bij enkele andere verdelingen.

In figuur 2.26 zie je een **tweetoppige verdeling**. Een verdeling waarbij een histogram hoort met meer dan één top heet een **meertoppige verdeling**.



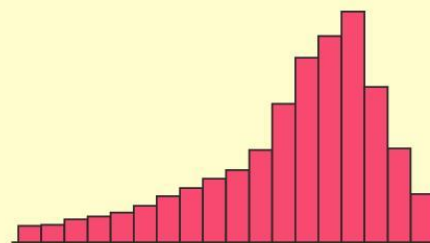
figuur 2.26

In figuur 2.27 zie je een **scheve verdeling**. Deze verdeling is **rechts-scheef** omdat het histogram aan de rechterkant erg ver doorloopt. Er is aan de rechterkant een staart bij het histogram.



figuur 2.27

In figuur 2.28 is een **links-scheve verdeling** getekend.



figuur 2.28

In figuur 2.29 ten slotte is een **uniforme verdeling** getekend. Dit is een verdeling waarbij elk waarnemingsgetal of elke klasse dezelfde frequentie heeft.



figuur 2.29

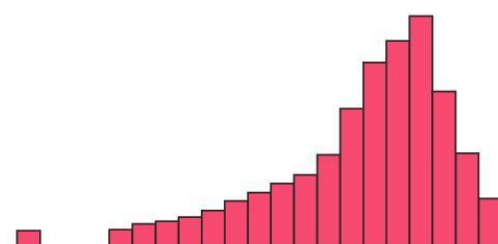
56



Zie de figuren 2.28 en 2.30. Het meest linkse staafje in figuur 2.28 is de uitschieter in figuur 2.30 geworden.

Vergelijk bij de figuren 2.28 en 2.30

- a het gemiddelde
- b de mediaan
- c de standaardafwijking.

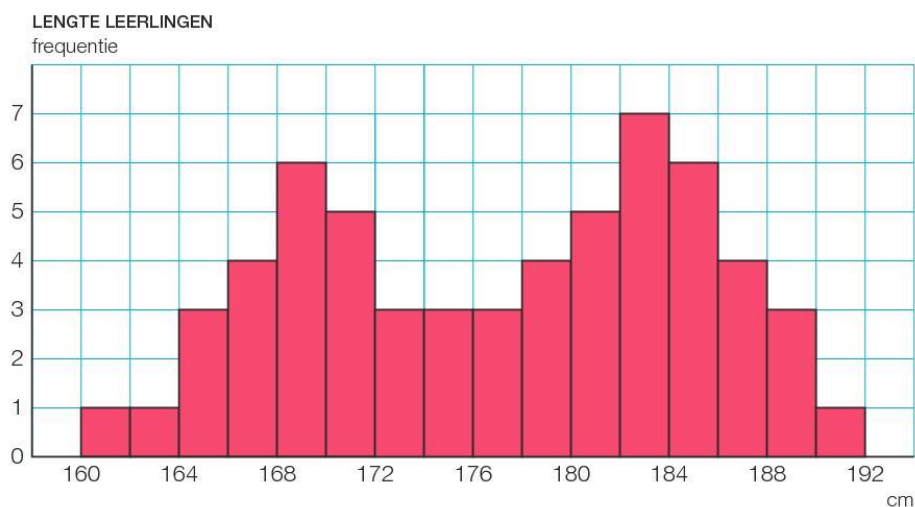


figuur 2.30

57

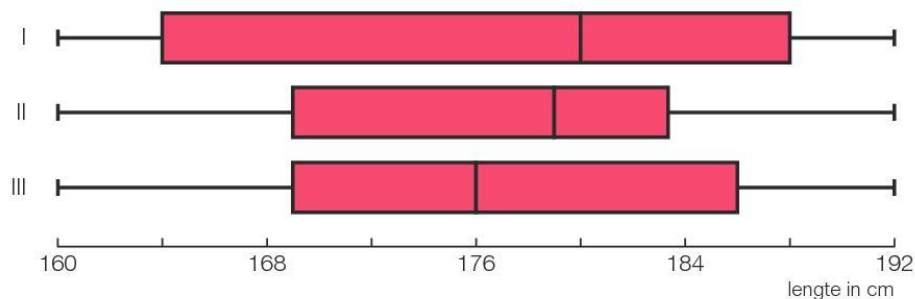



Zie figuur 2.31 met de verdeling van de lengten van 59 leerlingen. De eerste klasse is $[160, 162)$.



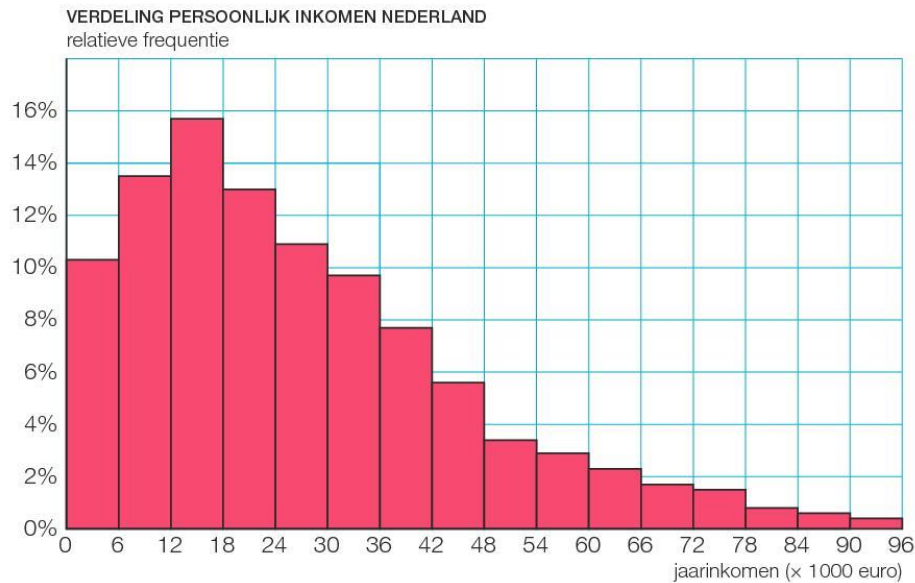
figuur 2.31

- Er zijn twee toppen in het histogram. Geef hiervoor een mogelijke verklaring.
- Geef de modale klasse.
- Geef een schatting van de standaardafwijking. Kies uit 1 cm, 8 cm, 15 cm en 30 cm.
- In de figuur hieronder zie je drie boxplots. Welke van de drie past het beste bij figuur 2.31? Licht toe.



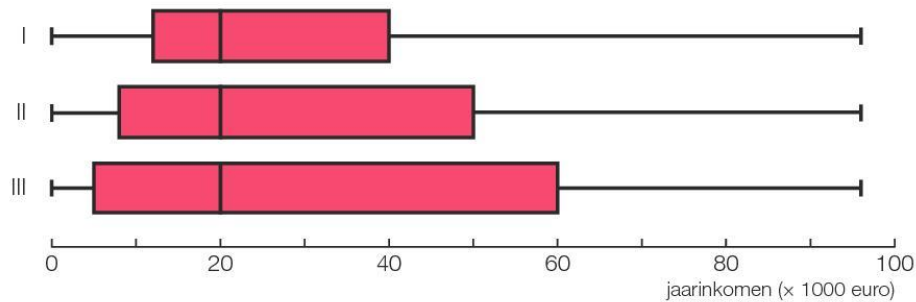
figuur 2.32

In figuur 2.33 zie je een gedeelte van de verdeling van het persoonlijk inkomen in Nederland.



figuur 2.33

- Zou de mediaan hier in de modale klasse kunnen liggen?
- Joost beweert dat als iedereen hetzelfde inkomen zou hebben, alle staven in figuur 2.33 even hoog zouden zijn. Dus dat je in dat geval met een uniforme verdeling te maken hebt. Heeft Joost gelijk? Licht toe.
- In de figuur hieronder zie je drie boxplots. Welke van de drie past het beste bij figuur 2.33? Licht toe.



figuur 2.34

59

Hieronder staan enkele situaties.

*

Kies bij elk van deze situaties welke verdeling erbij hoort.

Kies uit meertoppig, uniform, links-scheef, rechts-scheef en symmetrisch.

- a Het IQ van Nederlandse vrouwen van 20 jaar.
- b Het jaarinkomen van Nederlandse werknemers.
- c Het aantal uren gamen per week van leerlingen op een middelbare school.
- d Het gewicht van sinaasappels bij een kraam op de markt.
- e De cijfers van een moeilijke wiskundetoets.



A60

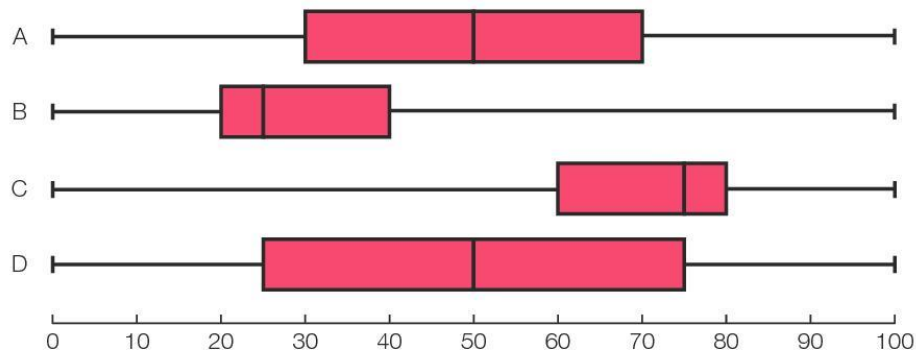


Van een grote serie waarnemingsgetallen is het kleinste waarnemingsgetal 0 en het grootste 100.

In figuur 2.35 zijn vier boxplots getekend. Deze horen bij de verdelingen I, II, III en IV.

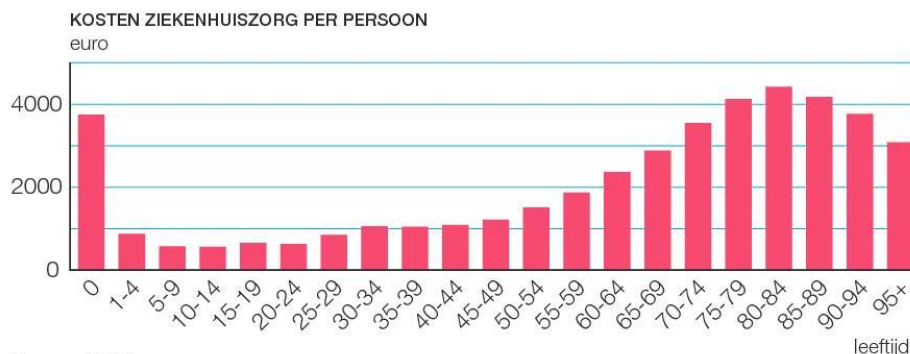
Welk boxplot hoort bij elk van deze verdelingen?

- I De verdeling is uniform.
- II De verdeling is tweetoppig, de eerste top zit bij 30 en de tweede top bij 70.
- III De verdeling is rechts-scheef.
- IV De verdeling is links-scheef.



figuur 2.35

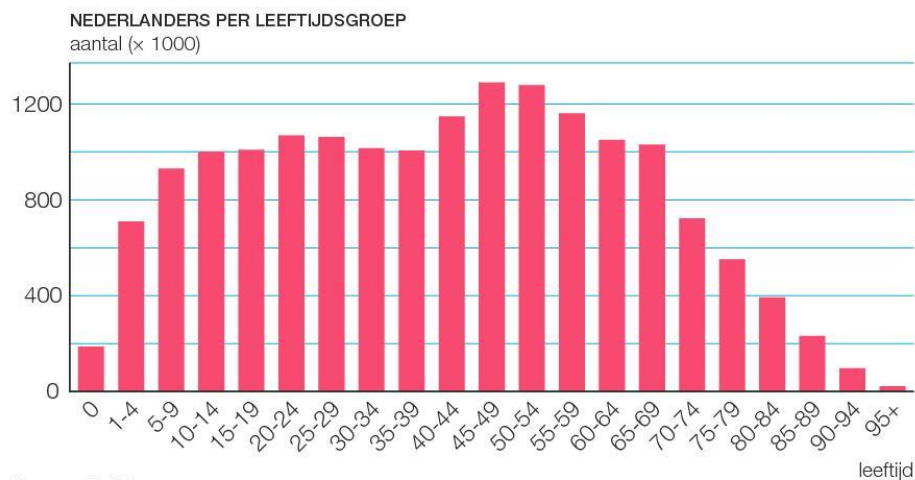
In figuur 2.36 zie je hoe de kosten van ziekenhuiszorg per persoon in Nederland zijn verdeeld naar leeftijdsgroep. De gegevens zijn uit 2015.



figuur 2.36

- De staaf bij de nuljarigen is veel hoger dan de staven ernaast. Geef hiervoor een verklaring.
- Gemiddeld werd per persoon 1593 euro uitgegeven. Vanaf welke leeftijdsgroep (uitgezonderd de nuljarigen) zijn de uitgaven per persoon meer dan gemiddeld?
- Irma beweert dat uit figuur 2.36 blijkt dat meer dan de helft van alle uitgaven voor ziekenhuiszorg voor rekening komt van de 70-plussers. Waarom kan Irma deze bewering alleen op grond van figuur 2.36 niet waarmaken?

In figuur 2.37 zie je de verdeling van de Nederlandse bevolking naar leeftijdsgroep in 2015.

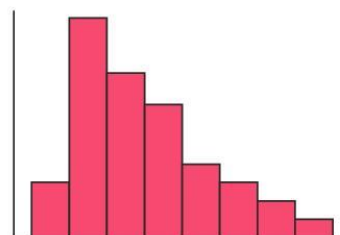


figuur 2.37

- De staaf bij de nuljarigen is veel lager dan de staven ernaast. Verklaar dit verschil.
- Met behulp van de figuren 2.36 en 2.37 kun je de totale ziekenhuiskosten per leeftijdsgroep berekenen. Onderzoek met een berekening welke leeftijdsgroep de hoogste totale kosten heeft, de groep van 45-49 jaar of de groep van 75-79 jaar.

Zie het histogram in figuur 2.38.

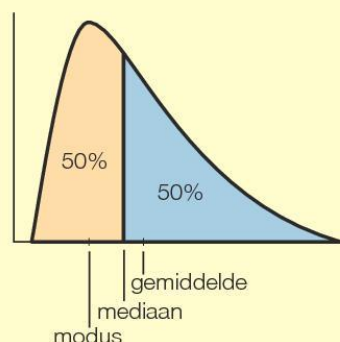
Is de modus groter of kleiner dan de mediaan? Licht toe.



figuur 2.38

Theorie B Verdelingskrommen

Een histogram geeft een goed beeld van een frequentieverdeling. Maar in plaats van een histogram te tekenen kun je ook een globale grafiek van de verdeling schetsen. Zo'n grafiek heet een **verdelingskromme**. Zo hoort bij het histogram in figuur 2.38 de verdelingskromme hiernaast. Ook hier ligt de modus bij de top. De verticale lijn bij de mediaan verdeelt de oppervlakte onder de kromme in twee gebieden met gelijke oppervlakte. De mediaan is hier dus groter dan de modus. Omdat de verdeling rechts-scheef is, is het gemiddelde groter dan de mediaan.



figuur 2.39

Ook de **cumulatieve verdelingskromme** geeft een beeld van de verdeling. Is een verdelingskromme gegeven dan kun je daarbij een schets maken van de cumulatieve verdelingskromme en andersom.

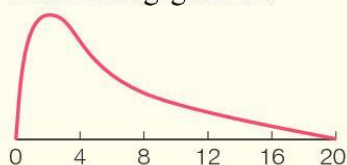
Voorbeeld

Zie de relatieve cumulatieve verdelingskromme in figuur 2.40.

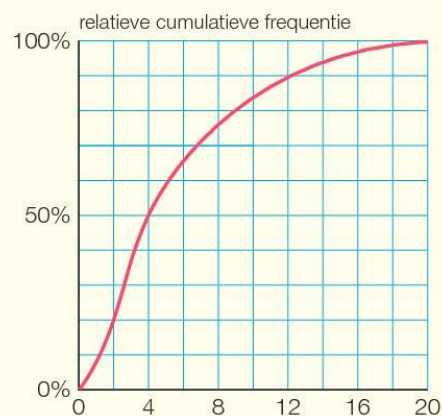
- Geef een schatting van de mediaan.
- Schets de bijbehorende verdelingskromme.
- Met wat voor soort verdeling heb je te maken?

Uitwerking

- Aflezen bij 50% geeft waarnemingsgetal 4. Dus de mediaan is 4.
- Je ziet in de gegeven figuur: in het begin is de kromme erg steil, dus zijn daar veel waarnemingsgetallen. Op het eind is de kromme minder steil, dus daar zijn weinig waarnemingsgetallen.



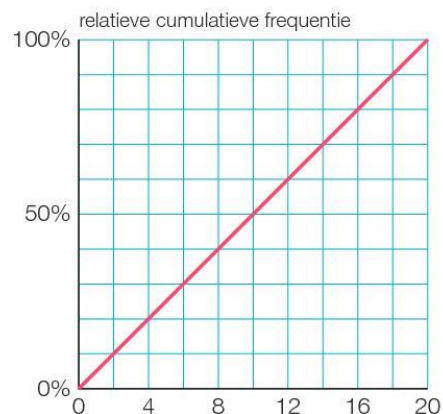
- Het is een rechts-scheve verdeling.



figuur 2.40

R63
□ ⊗ *

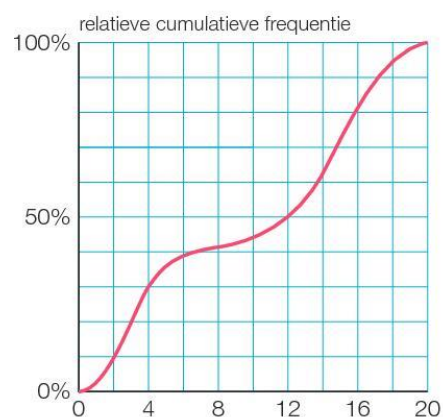
- a** Zie voorbeeld b.
Licht toe dat bij het steilste stuk van de relatieve cumulatieve verdelingskromme een top van de verdelingskromme hoort.
- b** Zie de relatieve cumulatieve verdelingskromme in figuur 2.41.
Met wat voor soort verdeling heb je te maken?



figuur 2.41

64
□ ⊗ *

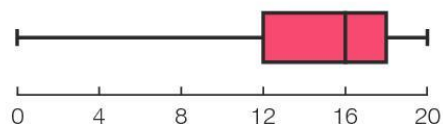
- Zie de relatieve cumulatieve verdelingskromme in figuur 2.42.
- a** Teken de bijbehorende boxplot.
- b** Schets de verdelingskromme die hoort bij figuur 2.42.
- c** Met wat voor soort verdeling heb je te maken?



figuur 2.42

65
□ ⊗ *

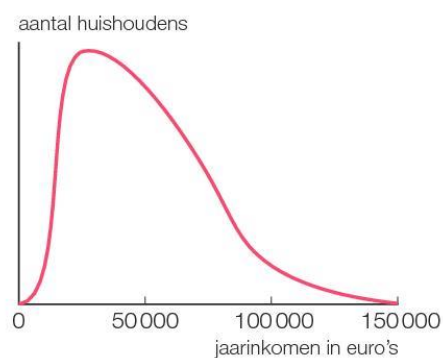
- Gegeven is de boxplot in figuur 2.43.
- a** Schets de bijbehorende verdelingskromme.
- b** Met wat voor soort verdeling heb je te maken?
- c** Hoe ligt het gemiddelde ten opzichte van de mediaan?



figuur 2.43

A66
□ ⊗ *

- De verdelingskromme in figuur 2.44 geeft informatie over de verdeling van de jaarinkomens van huishoudens in een land.
- a** Hoe ligt de mediaan ten opzichte van de modus bij deze verdeling?
- b** Schets de relatieve cumulatieve verdelingskromme.

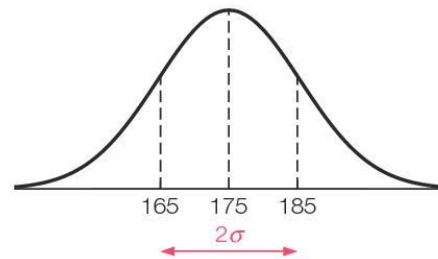


figuur 2.44

O67

In figuur 2.45 zie je een verdelingskromme bij een symmetrische verdeling. Het gemiddelde is 175 en de standaardafwijking $\sigma = 10$. Bij deze verdeling zit 68% van de waarnemingsgetallen tussen 165 en 185.

Hoeveel procent van de waarnemingsgetallen is groter dan 185?



figuur 2.45

Theorie C De normale verdeling

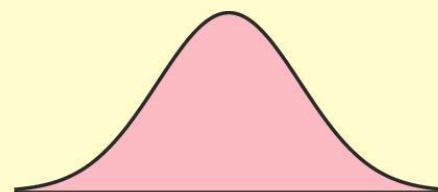
In figuur 2.46 is de verdelingskromme bij een bijzondere symmetrische verdeling getekend. Deze klokvormige kromme heet de **normaalkromme** en hoort bij de **normale verdeling**.

Bij de normale verdeling wordt het gemiddelde aangegeven met de Griekse letter μ . Spreek μ uit als mu.

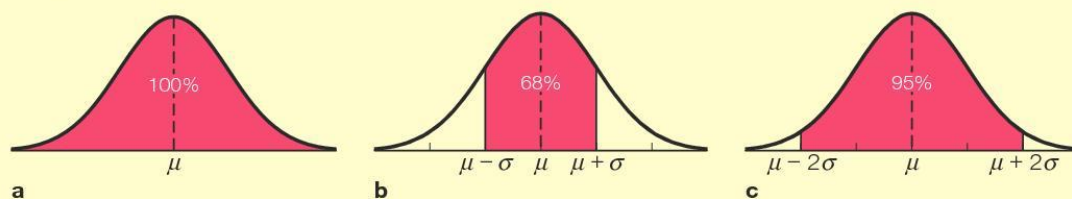
De totale oppervlakte onder de kromme is 100%.

De normale verdeling is zodanig gekozen dat ongeveer 68% van de waarnemingsgetallen minder dan σ van het gemiddelde μ afligt.

En ongeveer 95% van de waarnemingsgetallen ligt minder dan 2σ af van μ .



figuur 2.46 De normaalkromme



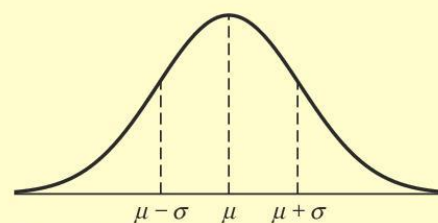
figuur 2.47

In figuur 2.48 is nogmaals een normaalkromme getekend.

Links van μ is de kromme stijgend. Je ziet dat de stijging links van $\mu - \sigma$ anders is dan de stijging tussen $\mu - \sigma$ en μ .

Links van $\mu - \sigma$ is de kromme **toenemend stijgend** en tussen $\mu - \sigma$ en μ is de kromme **afnemend stijgend**.

En zo is de kromme tussen μ en $\mu + \sigma$ **toenemend dalend** en rechts van $\mu + \sigma$ **afnemend dalend**.

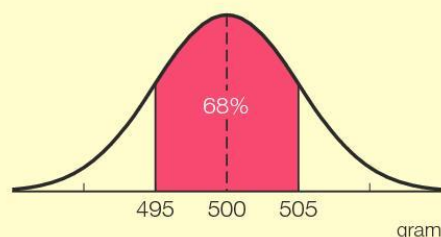


figuur 2.48

Normale verdelingen komen veel voor in het dagelijks leven. Voorbeelden zijn de lengte van de Nederlandse man, het aantal branduren van een grote partij lampen van 12 watt en de gewichten van een groot aantal pakken koffie.

Figuur 2.49 hoort bij de gewichten van pakken koffie waarvan de inhoud gemiddeld 500 gram is. De standaardafwijking is 5 gram. De inhoud is normaal verdeeld en daarom heeft ongeveer 68% van de pakken een inhoud tussen 495 en 505 gram.

Omdat $\mu - 2\sigma = 500 - 2 \cdot 5 = 490$ en $\mu + 2\sigma = 500 + 2 \cdot 5 = 510$, heeft ongeveer 95% van de pakken een inhoud tussen 490 en 510 gram.

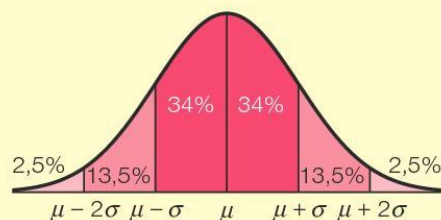


figuur 2.49

Uit de symmetrie van de normaalkromme volgen de percentages in de verschillende gebieden in figuur 2.50. Onthoud deze percentages.

Vuistregels bij de normale verdeling

- 68% van de waarnemingsgetallen ligt tussen $\mu - \sigma$ en $\mu + \sigma$
- 95% van de waarnemingsgetallen ligt tussen $\mu - 2\sigma$ en $\mu + 2\sigma$



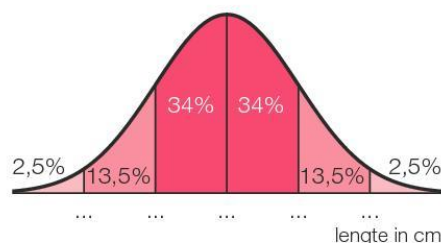
figuur 2.50

68 Van een groep volwassen vrouwen is de lengte normaal verdeeld met $\mu = 170$ cm en $\sigma = 5$ cm.

- a** Neem figuur 2.51 over en zet onder de horizontale as de juiste getallen.

Hoeveel procent van de vrouwen heeft volgens de vuistregels bij de normale verdeling een lengte

- b** tussen 165 en 180 cm
c minder dan 160 cm
d meer dan 175 cm
e tussen 160 en 170 cm?



figuur 2.51

69 In welk van de volgende gevallen is bij benadering sprake van een normale verdeling?

- a** Het jaarinkomen van de Nederlandse werknemers.
b De bloeddruk van de bejaarden in ons land.
c De leeftijd van de Nederlanders.
d De hoogte van de volgroeide eikenbomen op de Hoge Veluwe.
e De lengte van de telefoongesprekken op maandag.
f Het gewicht van kippeneieren uit een grote partij.

A70


Van 5000 appels is het gewicht normaal verdeeld met een gemiddelde van 180 gram en een standaardafwijking van 11 gram. Gebruik bij de volgende vragen de vuistregels bij de normale verdeling.

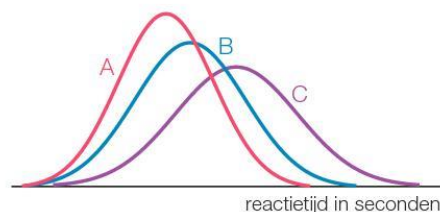
- Hoeveel van deze appels wegen tussen 158 en 191 gram?
- Hoeveel van deze appels zijn lichter dan 191 gram?
- Wat weet je van de gewichten van de zwaarste 125 appels van deze partij?

71


Bij drie grote groepen mannen wordt van elke man de reactietijd gemeten.

De eerste groep bestaat uit 18-jarigen, de tweede uit 30-jarigen en de derde uit 60-jarigen.

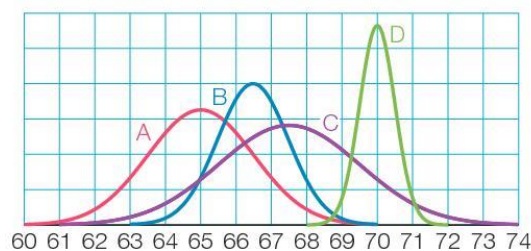
- Welke kromme hoort bij de 18-jarigen? En welke bij de 60-jarigen? Licht je antwoord toe.
- In welke groep is de kans het grootst dat de reactietijd van een persoon meer dan een halve seconde van het gemiddelde afwijkt?



figuur 2.52

A72


Zie de normaalkrommen in figuur 2.53. Geef bij elke kromme een schatting van het gemiddelde en van de standaardafwijking.



figuur 2.53

INFORMATIEF

De normale verdeling

Tallose beroemde geleerden hebben zich beziggehouden met de normale verdeling. Omstreeks 1730 gebruikte de Fransman Abraham de Moivre (1667-1754) als eerste de normale verdeling. Hij ontdekte dat vele kleine toevalligheden in de natuur merkwaardig genoeg tot mooie regelmatige verdelingen kunnen leiden. Ook de veelzijdige geleerde Gauss (1777-1855) werd gefascineerd door de normaalkromme en hij slaagde erin een formule op te stellen bij de kromme. Vandaar dat ook de naam Gauss-kromme gebruikt wordt. De Belg Adolphe Quetelet (1796-1874) ontdekte dat normaalkrommen goed gebruikt kunnen worden om menselijke kenmerken zoals lichaamslengte en voetlengte te beschrijven.



Carl Friedrich Gauss

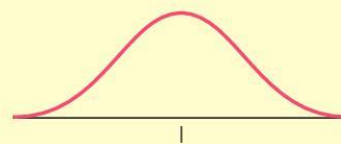
Terugblik

Verdelingen, centrummaten en spreiding

In de figuren hiernaast zie je voorbeelden van verdelingskrommen bij de volgende soorten verdelingen.

I symmetrische verdeling

Het gemiddelde, de modus en de mediaan vallen samen.



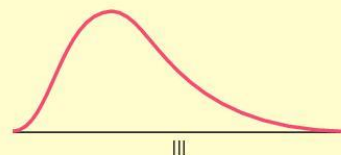
II tweetoppige verdeling

In de figuur hiernaast liggen het gemiddelde en de mediaan tussen de twee toppen.



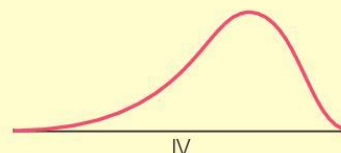
III rechts-scheve verdeling

Het gemiddelde en de mediaan liggen rechts van de top.



IV links-scheve verdeling

Het gemiddelde en de mediaan liggen links van de top.

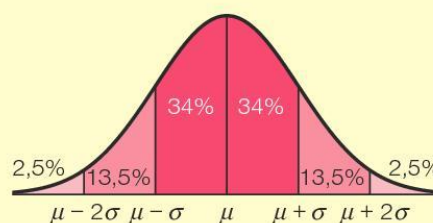
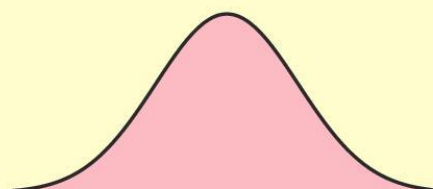


De normale verdeling

Heeft een symmetrische verdeling een klokvorm zoals hiernaast, dan is er sprake van een normale verdeling. De kromme heet normaalkromme. De top ligt boven het gemiddelde μ . De breedte van de kromme hangt af van de standaardafwijking σ .

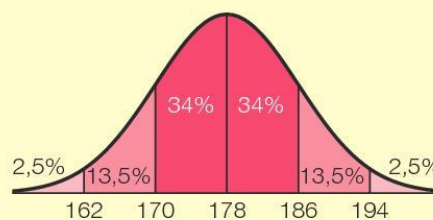
Bij een normale verdeling ligt

- 68% van alle waarnemingsgetallen tussen $\mu - \sigma$ en $\mu + \sigma$
- 95% van alle waarnemingsgetallen tussen $\mu - 2\sigma$ en $\mu + 2\sigma$.



Bij een groep mannen, waarvan de lengte normaal verdeeld is met $\mu = 178$ cm en $\sigma = 8$ cm, hoort de verdeling hiernaast.

De percentages volgen uit de vuistregels bij de normale verdeling. Je ziet dat 47,5% van de mannen een lengte heeft tussen 162 en 178 cm.



2.6 Betrouwbaarheidsintervallen

073



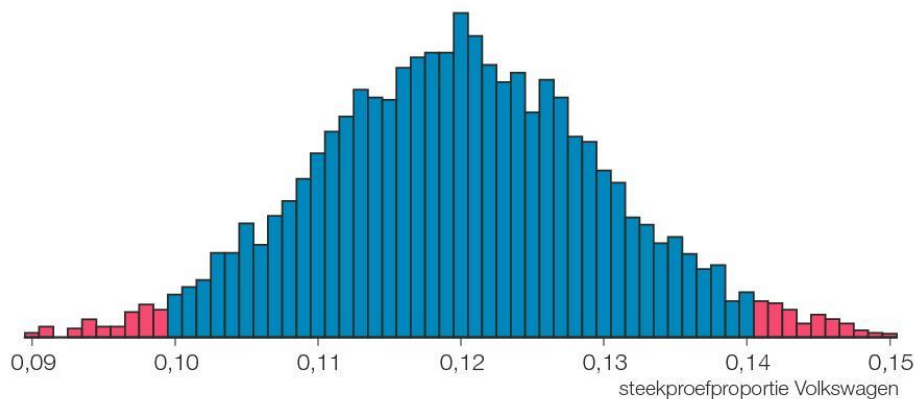
In deze opgave werk je met de module Steekproeven in Getal & Ruimte online.

In Nederland is 12% van het aantal personenauto's van het merk Volkswagen. Dus de populatieproportie van de Volkswagens is gelijk aan 0,12.

Noteer je van de eerste 100 personenauto's die je op een snelweg ziet het merk, dan is het echter niet gezegd dat hierbij precies 12 Volkswagens zijn. Het kunnen er ook meer of minder zijn.

- Gebruik een simulatie om een steekproef van lengte 100 bij een populatieproportie van 0,12 te nemen. Hoeveel Volkswagens vind je in de steekproef? Wat betekent dit voor de steekproefproportie?
- Simuleer nog enkele keren een steekproef van lengte 100 bij een populatieproportie van 0,12 en noteer telkens de steekproefproportie. Wat is de kleinste steekproefproportie in je simulatie? En de grootste?
- Simuleer enkele keren een steekproef van lengte 1000 bij een populatieproportie van 0,12 en noteer telkens de steekproefproportie. Wat is de kleinste steekproefproportie in je simulatie? En de grootste?
- Kun je op grond van de uitkomsten bij b en c iets zeggen over de invloed van de steekproeflengte op de grenzen waartussen de steekproefproporties liggen?

In figuur 2.54 zie je het histogram dat een resultaat geeft van 5000 steekproeven van lengte 1000 bij een populatieproportie van 0,12.



figuur 2.54

- Tussen welke grenzen ligt de steekproefproportie?
- In de figuur is ook aangegeven tussen welke grenzen 95% van de steekproefproporties ligt. Welke grenzen zijn dat?

Theorie A Betrouwbaarheidsintervallen voor populatieproportie

In deze paragraaf richten we ons op fase 4 van de statistische cyclus: Conclusies trekken.

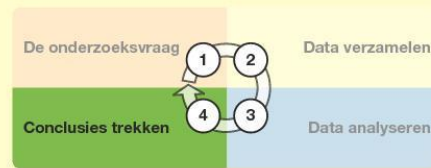
In figuur 2.53 van opgave 73 zie je dat het histogram lijkt op de normaalkromme. Door een groot aantal steekproeven met voldoende grote steekproeflengte te nemen zal elke steekproevenverdeling steeds meer gaan lijken op een normale verdeling.

Omdat in het model van de normale verdeling volgens een van de vuistregels 95% van alle waarnemingsgetallen ligt tussen $\mu - 2\sigma$ en $\mu + 2\sigma$ weet je nu ook dat de standaardafwijking van de steekproefresultaten gelijk is

$$\text{aan } \frac{0,14 - 0,10}{2} = 0,02.$$

Voor de standaardafwijking van een steekproevenverdeling bestaat de

formule $\sigma = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$. In deze formule is σ de standaardafwijking, $\hat{p}$ de steekproefproportie en n de steekproefomvang.



De lengte van de steekproef wordt steekproeflengte en ook wel steekproefomvang genoemd.

Bij een steekproevenverdeling is de steekproefproportie $\hat{p}$ bij een voldoende grote steekproefomvang n bij benadering normaal verdeeld. Hierbij is het gemiddelde μ gelijk aan de populatieproportie.

Voor de standaardafwijking σ geldt $\sigma = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$.

In opgave 73 gingen we ervan uit dat de populatieproportie bekend was en dat een flink aantal steekproeven werd genomen met een vaste steekproeflengte.

Deze situatie komt in de praktijk zelden voor. In de praktijk is juist vaak de populatieproportie niet bekend en wil men door middel van één steekproef hiervan een betrouwbare schatting krijgen.

Maar hoe groot moet deze steekproef zijn en hoe betrouwbaar is de schatting dan?

Als bij een onderzoek blijkt dat 60 van de 80 onderzochte tweelingen dezelfde favoriete sport te hebben, dan kunnen we niet zomaar zeggen dat 75% van alle tweelingen dezelfde favoriete sport heeft.

Omdat de standaardafwijking van de steekproevenverdeling gelijk is aan

$$\sigma = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0,75 \cdot 0,25}{80}} = 0,0484...$$

$$\text{is } 2\sigma = 2 \cdot 0,0484... \approx 0,097.$$

Hieruit volgt dat $\hat{p} - 2\sigma \approx 0,75 - 0,097 = 0,653$

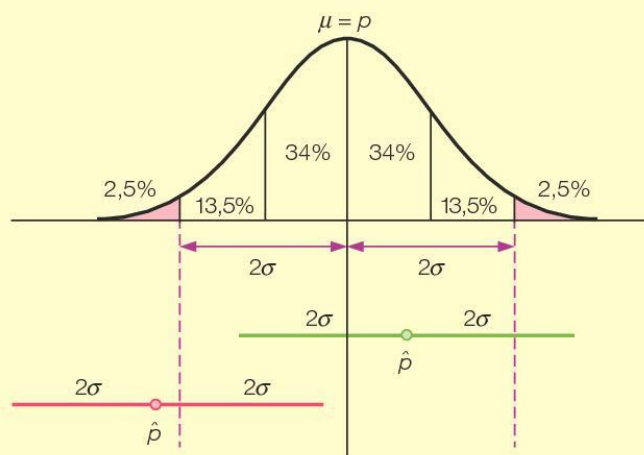
en $\hat{p} + 2\sigma \approx 0,75 + 0,097 = 0,847.$

We zeggen dat het **95%-betrouwbaarheidsinterval** $[0,653; 0,847]$ is. Dit wordt ook genoteerd als het 95%-betrouwbaarheidsinterval is $[65,3\%; 84,7\%]$.

De breedte van het 95%-betrouwbaarheidsinterval is 4σ .

Hieronder lichten we toe wat de precieze betekenis is van een 95%-betrouwbaarheidsinterval.

Neem je meerdere steekproeven uit een populatie, dan zullen de steekproefproporties niet steeds hetzelfde zijn. De meeste steekproefproporties zullen in de buurt van de populatieproportie liggen, maar er zullen er ook zijn die er ver vanaf liggen. De verdeling van deze steekproefproporties is bij benadering normaal verdeeld met $\mu = p$.



figuur 2.55

In figuur 2.55 zie je de normaalkromme bij deze verdeling, met daarin de percentages die je kent van de vuistregels bij de normale verdeling. Er zijn twee steekproefproporties $\hat{p}$ weergegeven. De lijnstukken waarvan de steekproefproporties het midden zijn, hebben een lengte van 4σ . Hierbij horen de intervallen $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$. Het groene interval bevat de populatieproportie p en het rode bevat p niet.

Krijg je bij een steekproef een $\hat{p}$ die tussen de stippellijnen ligt, dan ligt deze steekproefproportie minder dan 2σ van p af. In dat geval zal het interval $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$ de populatieproportie p bevatten. Uit een van de vuistregels voor de normale verdeling volgt dat dit in 95% van de steekproeven het geval is.

Neem je één steekproef, dan weet je niet zeker of het bij deze steekproef opgestelde interval $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$ de populatieproportie p bevat. Je weet immers niet zeker of het verkregen interval hoort bij de 95% die p wel bevat of bij de 5% die p niet bevat. Daarom wordt het interval het 95%-betrouwbaarheidsinterval genoemd.

Voor een populatieproportie is het 95%-betrouwbaarheidsinterval gelijk aan $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$. Hierin is $\hat{p}$ de steekproefproportie en σ de standaardafwijking.

Het 95%-betrouwbaarheidsinterval geeft aan dat bij herhaling van de procedure met steeds een nieuwe aselechte steekproef uit dezelfde populatie, 95% van de daarbij opgestelde intervallen de populatieproportie bevat.

Omdat de breedte van een betrouwbaarheidsinterval afhangt van de standaardafwijking $\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, zal de breedte van een betrouwbaarheidsinterval afnemen bij een grotere steekproefomvang n . Immers, hoe groter n , hoe kleiner σ en dus hoe smaller het betrouwbaarheidsinterval.

Heb je bijvoorbeeld te maken met een steekproefproportie $\hat{p} = 0,25$, dan is bij een steekproefomvang $n = 100$ de standaardafwijking

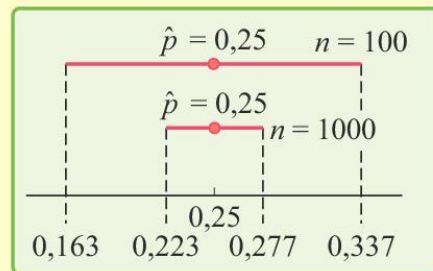
$$\sigma = \sqrt{\frac{0,25 \cdot 0,75}{100}} = 0,0433... \text{ Dus het}$$

95%-betrouwbaarheidsinterval is $[0,25 - 2 \cdot 0,0433...; 0,25 + 2 \cdot 0,0433...]$ oftewel ongeveer $[0,163; 0,337]$.

Neem je echter steekproefomvang $n = 1000$,

$$\text{dan is } \sigma = \sqrt{\frac{0,25 \cdot 0,75}{1000}} = 0,0136... \text{ en is het}$$

95%-betrouwbaarheidsinterval ongeveer $[0,223; 0,277]$.



Voorbeeld

Twee dagen voor de verkiezingen voor de Tweede Kamer wordt aan 1200 stemgerechtigde Nederlanders gevraagd of ze van plan zijn te gaan stemmen. Van deze 1200 zeggen 756 inderdaad te gaan stemmen.

Bereken het 95%-betrouwbaarheidsinterval voor de proportie stemgerechtigde Nederlanders die van plan zijn te gaan stemmen.

Uitwerking

$$\hat{p} = \frac{756}{1200} = 0,63 \text{ en } \sigma = \sqrt{\frac{0,63 \cdot 0,37}{1200}} = 0,0139...$$

$$\hat{p} - 2\sigma = 0,63 - 2 \cdot 0,0139... \approx 0,602 \text{ en}$$

$$\hat{p} + 2\sigma = 0,63 + 2 \cdot 0,0139... \approx 0,658$$

Het 95%-betrouwbaarheidsinterval is $[0,602; 0,658]$.

Rond in een betrouwbaarheidsinterval af op drie decimalen, tenzij anders wordt gevraagd.

74

Zie het voorbeeld.



Bij een andere peiling worden 600 stemgerechtigde Nederlanders ondervraagd. Hiervan zijn er 378 van plan te gaan stemmen.

- a Toon aan dat de steekproefproportie ook nu gelijk is aan 0,63 en bereken de standaardafwijking. Rond af op vier decimalen.
- b Bereken het 95%-betrouwbaarheidsinterval voor de proportie stemgerechtigde Nederlanders die van plan zijn te gaan stemmen.
- c Vergelijk het 95%-betrouwbaarheidsinterval van vraag b met dat van het voorbeeld.

Welke conclusie trek je?

75



Je kunt ook een betrouwbaarheidsinterval opstellen met een andere betrouwbaarheid dan 95%. Met de kennis van dit hoofdstuk kun je bijvoorbeeld een 68%-betrouwbaarheidsinterval opstellen. Bedenk hoe dat moet en stel het 68%-betrouwbaarheidsinterval op bij het voorbeeld.

76



Een bioloog onderzoekt de verhouding tussen mannetjes en vrouwtjes van de bruine kikker in de bergmeren van de Pyreneeën.

Van de 150 onderzochte kikkers blijken er maar liefst 117 een mannetje te zijn.

- a Bereken de steekproefproportie van de mannetjeskikkers.
- b Bereken het 95%-betrouwbaarheidsinterval voor de proportie mannetjeskikkers.



77



Jochem doet een onderzoek naar de kleur van de personenauto's die in Nederland rondrijden. Hij noteert van 575 auto's de kleur. Hiervan zijn er 47 rood.

- a Bereken het 95%-betrouwbaarheidsinterval voor de proportie rode auto's.
- b Voor de proportie zwarte auto's is het 95%-betrouwbaarheidsinterval $[0,224; 0,296]$.

Bereken de standaardafwijking van het aantal zwarte auto's in de steekproef.

A78



Bij de presidentsverkiezingen in een land moet gekozen worden tussen de kandidaten A en B. In een peiling die voorafgaat aan de verkiezingen zeggen 935 mensen te gaan stemmen. Hiervan zeggen 487 op kandidaat A te gaan stemmen, de anderen kiezen voor kandidaat B. Onderzoek of je met een betrouwbaarheid van 95% kunt zeggen dat kandidaat A de verkiezingen zal gaan winnen.

A79
☐ ⊗ *

Lowlands is een driedaags muziekfestival dat jaarlijks wordt gehouden in Biddinghuizen (Flevoland). In 2014 waren er 55 000 bezoekers.

Bij een onderzoek worden 1250 festivalgangers ondervraagd.

Het blijkt dat hiervan 225 in de provincie Zuid-Holland wonen.

- a** Bereken het 95%-betrouwbaarheidsinterval voor de proportie bezoekers die in Zuid-Holland woont.



Bij het onderzoek waren 715 mannen en 535 vrouwen betrokken. Van de mannen woonden er 143 in de provincie Zuid-Holland.

- b** Bereken voor de mannen het 95%-betrouwbaarheidsinterval voor de proportie die in Zuid-Holland woont.
- c** Bereken voor de vrouwen het 95%-betrouwbaarheidsinterval voor de proportie die in Zuid-Holland woont.
- d** Zet de drie betrouwbaarheidsintervallen van a, b en c op volgorde van breedte van het interval.
Hoe kun je de verschillende breedtes verklaren?

Bij het onderzoek is ook gevraagd naar de leeftijd van de bezoekers. Het 95%-betrouwbaarheidsinterval voor de proportie bezoekers met een leeftijd van 18-24 jaar was $[0,216; 0,264]$.

- e** Bereken de standaardafwijking van de proportie bezoekers met een leeftijd van 18-24 jaar.
- f** Bereken het aantal bezoekers in de steekproef met een leeftijd van 18-24 jaar.
- g** Bereken het 95%-betrouwbaarheidsinterval voor het aantal bezoekers met een leeftijd van 18-24 jaar.

A80
*

Naar aanleiding van de gemeenteraadsverkiezingen in de gemeente Lochem in 2018 heeft een lokaal dagblad een opiniepeiling gehouden onder de stemgerechtigden in Lochem. Het bijbehorend 95%-betrouwbaarheidsinterval voor het percentage kiezers dat voor D66 zegt te gaan stemmen is $[14,6\%; 19,4\%]$.

- a** Bereken het aantal stemgerechtigden in de steekproef van de opiniepeiling.

Na afloop van de verkiezingen blijkt D66 slechts 8,7% van de stemmen te hebben gekregen.

- b** Noem een mogelijke oorzaak voor de grote afwijking ten opzichte van het betrouwbaarheidsinterval.



Het gemiddelde geboortegewicht van Nederlandse jongens is normaal verdeeld met een gemiddelde van 3750 gram. De bijbehorende standaardafwijking is 480 gram.

- a In 2017 zijn in Nederland 87 159 jongens geboren.
Hoeveel van deze jongens waren naar verwachting zwaarder dan 4710 gram? Rond af op tientallen.
- b In 2017 werden in Oost-Nederland 18 138 jongens geboren.
Wat weet je van de gewichten van ongeveer 12 332 van deze jongens waarvan het gewicht het minst afwijkt van het gemiddelde gewicht?

Theorie B Betrouwbaarheidsintervallen voor populatiegemiddelde

Voor een onderzoek zijn allerlei gegevens verzameld van 36 pasgeboren jongens die na een normale zwangerschapsduur zijn geboren. Het gemiddelde geboortegewicht van deze jongens is 3722 gram en de standaardafwijking is 516 gram. Met deze gegevens is met 95% betrouwbaarheid een schatting te geven van het gemiddelde geboortegewicht van alle jongens die na een normale zwangerschapsduur zijn geboren.

Bij het 95%-betrouwbaarheidsinterval voor het **populatiegemiddelde** gebruiken we het **steekproefgemiddelde** $\bar{X}$, de **steekproefstandaardafwijking** S en de steekproefomvang n .

Het 95%-betrouwbaarheidsinterval is

$$\left[\bar{X} - 2 \cdot \frac{S}{\sqrt{n}}, \bar{X} + 2 \cdot \frac{S}{\sqrt{n}} \right].$$

Omdat bij de geboortegewichten $\bar{X} = 3722$, $S = 516$ en $n = 36$ is het 95%-betrouwbaarheidsinterval

$$\left[3722 - 2 \cdot \frac{516}{\sqrt{36}}, 3722 + 2 \cdot \frac{516}{\sqrt{36}} \right] = [3550, 3894].$$

Het 95%-betrouwbaarheidsinterval geeft aan dat met 95% zekerheid het gemiddelde geboortegewicht van de jongens tussen 3550 gram en 3894 gram ligt.

Er is verschil tussen de populatiestandaardafwijking σ en de steekproefstandaardafwijking S . Zie het informatief op de volgende bladzijde.

Voor een populatiegemiddelde is het 95%-betrouwbaarheidsinterval

gelijk aan $\left[\bar{X} - 2 \cdot \frac{S}{\sqrt{n}}, \bar{X} + 2 \cdot \frac{S}{\sqrt{n}} \right]$.

Hierin is $\bar{X}$ het steekproefgemiddelde, S de steekproefstandaardafwijking en n de steekproefomvang.

Het 95%-betrouwbaarheidsinterval geeft aan dat het interval met 95% zekerheid het populatiegemiddelde bevat.

Voorbeeld

Van 40 meisjes van 11 maanden is het gemiddelde gewicht 9180 gram en de standaardafwijking 1060 gram.

Bereken het 95%-betrouwbaarheidsinterval voor het gewicht van 11 maanden oude meisjes en geef aan welke informatie dat in deze context geeft.

Uitwerking

Het 95%-betrouwbaarheidsinterval is

$$\left[9180 - 2 \cdot \frac{1060}{\sqrt{40}}, 9180 + 2 \cdot \frac{1060}{\sqrt{40}} \right] \approx [8845, 9515].$$

Met 95% zekerheid zal het gemiddelde gewicht van 11 maanden oude meisjes liggen tussen 8845 gram en 9515 gram.

INFORMATIEF

σ versus S

In de statistiek worden twee soorten standaardafwijkingen gebruikt: de populatiestandaardafwijking σ en de steekproefstandaardafwijking S . Welke je gebruikt hangt af van de situatie.

Wanneer je iets wil zeggen over de gehele populatie en de populatiestandaardafwijking σ is beschikbaar, dan gebruik je deze.

Soms is de populatiestandaardafwijking niet bekend. Als er in dat geval een steekproef wordt genomen, dan kun je de steekproefstandaardafwijking berekenen. Deze gebruik je dan om iets over de hele populatie te zeggen.

Hoe groter de steekproef, hoe kleiner het verschil is tussen σ en S .

In de opgaven gebruik je de gegeven standaardafwijking.

82



Bij een steekproef van 65 basisschoolleerlingen met zwemdiploma A is de gemiddelde leeftijd waarop zij het diploma hebben gehaald gelijk aan 6 jaar en 9 maanden. De standaardafwijking is 10 maanden. Bereken in tienden van maanden nauwkeurig het 95%-betrouwbaarheidsinterval voor de gemiddelde leeftijd waarop basisschoolleerlingen het zwemdiploma A halen. Welke informatie geeft dat in deze context?



- 83** Bij een steekproef van lengte 35 is het 95% betrouwbaarheidsinterval voor het gemiddelde $[170,2; 179,8]$.
Bereken het steekproefgemiddelde en de steekproefstandaardafwijking.

- A84** Van 39 meisjes en 43 jongens is bij een leeftijd van 6 maanden en 12 maanden de lengte genoteerd. Zie de tabel.

	aantal	lengte in cm			
		6 maanden		12 maanden	
		gem.	stand. afw.	gem.	stand. afw.
meisjes	38	66,2	2,8	73,6	3,2
jongens	43	68,0	3,1	76,8	3,4

Bij deze tabel kunnen vier 95%-betrouwbaarheidsintervallen worden opgesteld.
Onderzoek welk van de vier intervallen het smalst is.



- E85** Bij een steekproef waarbij de steekproefstandaardafwijking 10 is, is het 95%-betrouwbaarheidsinterval voor het gemiddelde $[178, 182]$.
Bereken de steekproefomvang.

Terugblik

Gemiddelde en standaardafwijking van proporties

Bij een steekproevenverdeling is de steekproefproportie $\hat{p}$ bij een voldoende grote steekproefomvang n bij benadering normaal verdeeld. Hierbij is het gemiddelde μ gelijk aan de populatieproportie. Voor de

standaardafwijking geldt $\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Betrouwbaarheidsintervallen voor populatieproportie

Voor een populatieproportie is het 95%-betrouwbaarheidsinterval gelijk aan $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$.

Bij een onderzoek naar de verkeersveiligheid van scooters zijn 760 volwassenen ondervraagd. Hiervan zijn 213 voor een helmplicht voor scooterrijders.

Bij dit onderzoek bereken je het 95%-betrouwbaarheidsinterval voor de proportie volwassenen die voor een helmplicht voor scooterrijders zijn als volgt.

$$\hat{p} = \frac{213}{760} = 0,280... \text{ en } \sigma = \sqrt{\frac{0,280... \cdot 0,719...}{760}} = 0,016...$$

$$\hat{p} - 2\sigma = 0,280... - 2 \cdot 0,016... \approx 0,248 \text{ en}$$

$$\hat{p} + 2\sigma = 0,280... + 2 \cdot 0,016... \approx 0,313$$

Het 95%-betrouwbaarheidsinterval is $[0,248; 0,313]$.

Dus met 95% zekerheid zal tussen 24,8% en 31,3% van de volwassenen voor een helmplicht voor scooterrijders zijn.

Betrouwbaarheidsintervallen voor populatiegemiddelde

Voor een populatiegemiddelde is het 95%-betrouwbaarheidsinterval

gelijk aan $\left[\bar{X} - 2 \cdot \frac{S}{\sqrt{n}}, \bar{X} + 2 \cdot \frac{S}{\sqrt{n}} \right]$. Hierin is $\bar{X}$ het

steekproefgemiddelde, S de steekproefstandaardafwijking en n de steekproefomvang.

Van 40 jongens van 7 maanden is het gewicht bepaald. Het gemiddelde is 8,1 kg en de standaardafwijking 0,9 kg.

$$\text{Dus } \frac{S}{\sqrt{n}} = \frac{0,9}{\sqrt{40}} = 0,142... \text{ en dus } \bar{X} - 2 \cdot \frac{S}{\sqrt{n}} = 8,1 - 2 \cdot 0,142... \approx 7,82 \text{ en}$$

$$\bar{X} + 2 \cdot \frac{S}{\sqrt{n}} = 8,1 + 2 \cdot 0,142... \approx 8,38.$$

Het 95%-betrouwbaarheidsinterval voor het gemiddelde gewicht in kg van 7 maanden oude jongens is $[7,82; 8,38]$.

Eindopdracht Welbevinden op school

Het HSBC-onderzoek over gezondheid en welzijn van Nederlandse jongeren wordt elke vier jaar uitgevoerd. In het rapport worden niet alleen de gegevens van de laatste jaargang gepubliceerd, maar wordt ook gekeken naar hoe de gegevens zich in de tijd ontwikkelen. In onderstaande tabel zie je gegevens over jongens en meisjes in het voorgezet onderwijs van 2001 tot en met 2017.

De gegevens in de tabel zijn percentages.

	jongens					meisjes				
jaartal	'01	'05	'09	'13	'17	'01	'05	'09	'13	'17
vindt school leuk	80,4	80,3	79,6	79,9	73,9	82,3	84,8	86,2	84,6	79,0
voelt veel druk door schoolwerk	12,8	16,3	17,4	22,9	29,3	19,0	22,1	24,1	32,6	42,0
spijbelt wel eens	-	15,0	14,4	10,4	13,1	-	13,6	9,4	8,2	11,8
vindt sfeer onder klasgenoten goed	75,6	84,5	84,8	88,8	84,3	77,5	83,9	87,6	86,8	83,4
vindt relatie met leraren goed	-	-	-	68,2	65,3	-	-	-	66,2	63,2
heeft vaak gepest	16,2	12,3	9,4	5,9	2,9	6,2	5,8	4,2	3,1	1,1
wordt vaak gepest	10,6	8,1	8,0	7,0	4,7	8,3	6,0	4,6	6,9	4,2

In de figuur hiernaast zie je hoeveel jongens en meisjes elk jaar zijn ondervraagd. We noemen verschillen tussen bijvoorbeeld twee jaartallen of tussen jongens en meisjes significant als de 95%-betrouwbaarheidsintervallen elkaar niet overlappen. Onderzoek met behulp van de tabel of er een significant verschil bestaat tussen

- jongens en meisjes in 2017 als het gaat om *voelt veel druk door schoolwerk*
- jongens die vaak worden gepest in de jaren 2009 en 2017
- jongens en meisjes samen in 2013 en jongens en meisjes samen in 2017 als het gaat om spijbelen.



Diagnostische toets

2.1 De statistische cyclus

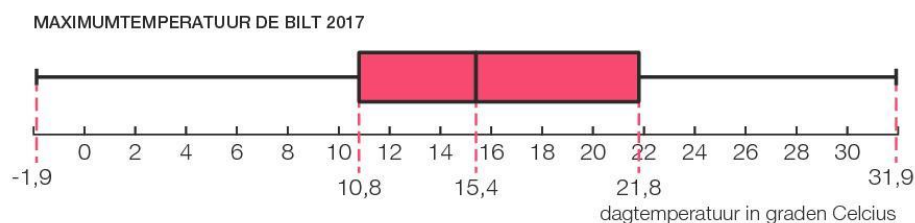
- 1 a Welke vier fasen onderscheiden we in de statistische cyclus?
b Wat wordt bedoeld met een causaal verband? Geef een voorbeeld.
c Aan welke twee eisen moet een representatieve steekproef voldoen?

2.2 Data verzamelen

- 2 a Omschrijf het verschil tussen een kwantitatieve en een kwalitatieve variabele. Geef van elk een voorbeeld.
b Omschrijf het verschil tussen een discrete en een continue variabele. Geef van elk een voorbeeld.
c Welke vier meetniveaus zijn er? Geef bij elk van de meetniveaus een variabele die op dat niveau is gemeten.
d Wat is het verschil tussen de populatieproportie en de steekproefproportie? Wanneer zullen deze proporties weinig verschillen?

2.3 Centrum- en spreidingsmaten

- 3 In het jaar 2017 is elke dag de maximumtemperatuur in De Bilt genoteerd. Zie de boxplot in figuur 2.56.



figuur 2.56

- a Hoeveel dagen is in 2017 de maximumtemperatuur meer dan $21,8^{\circ}\text{C}$ geweest?
- b Onderzoek of de gemiddelde maximumtemperatuur meer dan 20°C kan zijn geweest.
- c In de Bilt noteert men elke dag de maximumtemperatuur waarbij men afrondt op één decimaal. De modus is $15,2^{\circ}\text{C}$. Hoe vaak is minimaal $15,2^{\circ}\text{C}$ genoteerd? Licht toe.
- d Volgens Lars is de spreidingsbreedte meer dan drie keer zo groot als de interkwartielafstand. Onderzoek of Lars gelijk heeft.
- e Geef een schatting van de standaardafwijking. Kies uit 1°C , 4°C , 9°C en 14°C .

2.4 Klassenindeling en polygonen

- 4** In een stad is gedurende een jaar bijgehouden hoeveel klachten over geluidsoverlast er per dag binnenkwamen. Zie de frequentietabel.

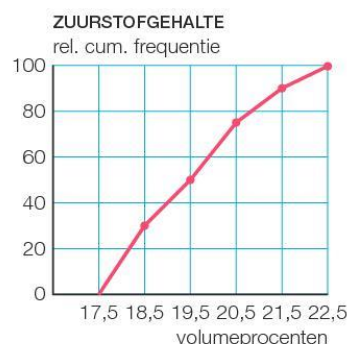
- a** Een jaar eerder waren er 6215 klachten over geluidsoverlast. Kun je op basis van de tabel concluderen of er dit jaar meer klachten waren? Licht toe.
- b** In welke klasse ligt de mediaan?
- c** Geef de modale klasse.

KLACHTEN GELUIDSOVERLAST

aantal	frequentie
[0, 5>	2
[5, 10>	6
[10, 15>	55
[15, 20>	126
[20, 25>	135
[25, 30>	35
[30, 35>	6

- 5** Bij een medisch onderzoek meet men bij 240 mannen het zuurstofgehalte (in volumeprocenten) van het bloed. Zie de figuur hiernaast, waarbij de eerste klasse [17,5; 18,5> is.

- a** Hoeveel mannen hebben minder dan 18,5% zuurstof in het bloed? En hoeveel mannen minstens 20,5%?
- b** Maak bij de figuur een frequentieverdeling met absolute frequenties.
- c** Teken een histogram met absolute frequenties.



figuur 2.57

2.5 Soorten verdelingen

- 6** Zie de boxplot in figuur 2.56. Schets bij deze figuur de verdelingskromme.
- 7** De inhoud van 750 potten pindakaas is normaal verdeeld met een gemiddeld gewicht van 360 gram en een standaardafwijking van 5 gram. Hoeveel van deze potten hebben volgens de vuistregels bij de normale verdeling een inhoud
- a** van meer dan 350 gram
 - b** tussen 365 gram en 370 gram?

2.6 Betrouwbaarheidsintervallen

- 8** Bij een onderzoek naar de financiële situatie van studenten zijn 500 studenten ondervraagd. Van deze studenten hadden er 80 een betalingsachterstand.
- a** Bereken het 95%-betrouwbaarheidsinterval voor de proportie studenten met een betalingsachterstand.

De gemiddelde betalingsachterstand van deze 80 studenten is 350 euro. De standaardafwijking is 108 euro.

- b** Bereken het 95%-betrouwbaarheidsinterval voor de gemiddelde betalingsachterstand van studenten met een betalingsachterstand. Rond af op gehele euro's.