

& GETAL & RUIMTE

A



Noordhoff Uitgevers

Legenda

1 Voorkennis

Kennis van enkele onderwerpen uit de onderbouw of uit een voorgaand hoofdstuk die je paraat moet hebben.

02 Oriëntatie-opgave

Opgave waarmee je je oriënteert op de theorie erna.

3 Gewone opgave

Na de theorie ga je oefenen met de gewone opgaven.

R4 Reflectie-opgave

In een reflectie-opgave kijk je nog eens terug op een voorgaand probleem.

A5 Afsluitende opgave

De afsluitende opgaven geven het beoogde beheersingsniveau aan.

E6 Extra opgave

Opgaven waarmee je extra wordt uitgedaagd.

Route-aanduiding

1 Basisroute

2 Middenroute

3 Uitdagende route

NB De symbolen van de route-aanduiding kunnen in combinaties voorkomen.

1 Opgaven zonder route-aanduiding

In de Diagnostische toets en in de Gemengde opgaven hebben opgaven geen route-aanduiding.

[► GR] Verwijzing naar een module van de GR-handleiding.

[► WERKBLAD] Verwijzing naar een werkblad.

Inhoud

1	Tabellen en grafieken	6	3	Lineaire verbanden	104
	Beginopdracht De Triathlon van Maastricht	8		Beginopdracht Draf en galop	106
	Voorkennis Procenten	9		Voorkennis Haakjes wegwerken en vergelijkingen oplossen	107
1.1	Rekenen met procenten en verhoudingen	11	3.1	Lineaire formules	109
1.2	Maatsystemen	22	3.2	Lineaire formules vergelijken	117
1.3	Omgaan met tabellen	31	3.3	Lineaire formules opstellen	126
1.4	Informatie in grafieken	41	3.4	Lineaire vergelijkingen met twee variabelen	136
	Eindopdracht RUN Winschoten	49		Eindopdracht Man en paard	141
	Diagnostische toets	50		Diagnostische toets	142
2	De statistische cyclus	52	4	Handig tellen	144
	Beginopdracht Elektrisch rijden	54		Beginopdracht Spelbord ontwerp	146
	Voorkennis Centrummaten en frequentietabellen	55		Voorkennis Teldiagrammen	147
2.1	De onderzoeksvraag	57	4.1	Regels bij telproblemen	148
2.2	Data verzamelen	66	4.2	Permutaties en combinaties	156
2.3	Centrum- en spreidingsmaten	72	4.3	Allerlei telproblemen	165
2.4	Representatie van data	80		Eindopdracht Cirkelvormig spelbord	177
2.5	Conclusies trekken	91		Diagnostische toets	178
	Eindopdracht Nederlands energieverbruik	101			
	Diagnostische toets	102		Gemengde opgaven	180
				Formuleblad	203
				Overzicht GR-modules	204
				Overzicht routes	205
				Trefwoordenregister	209
				Verantwoording	211



2

De statistische cyclus

Wat leer je?

- Welke fasen worden onderscheiden in de statistische cyclus.
- Aan welke voorwaarden een representatieve steekproef moet voldoen.
- Wat statistische variabelen zijn en hoe je ze kunt indelen.
- Diagrammen analyseren en in elkaar omzetten.
- De voor- en nadelen van het werken met een klassenindeling.
- Het trekken van conclusies met behulp van associatiematen.

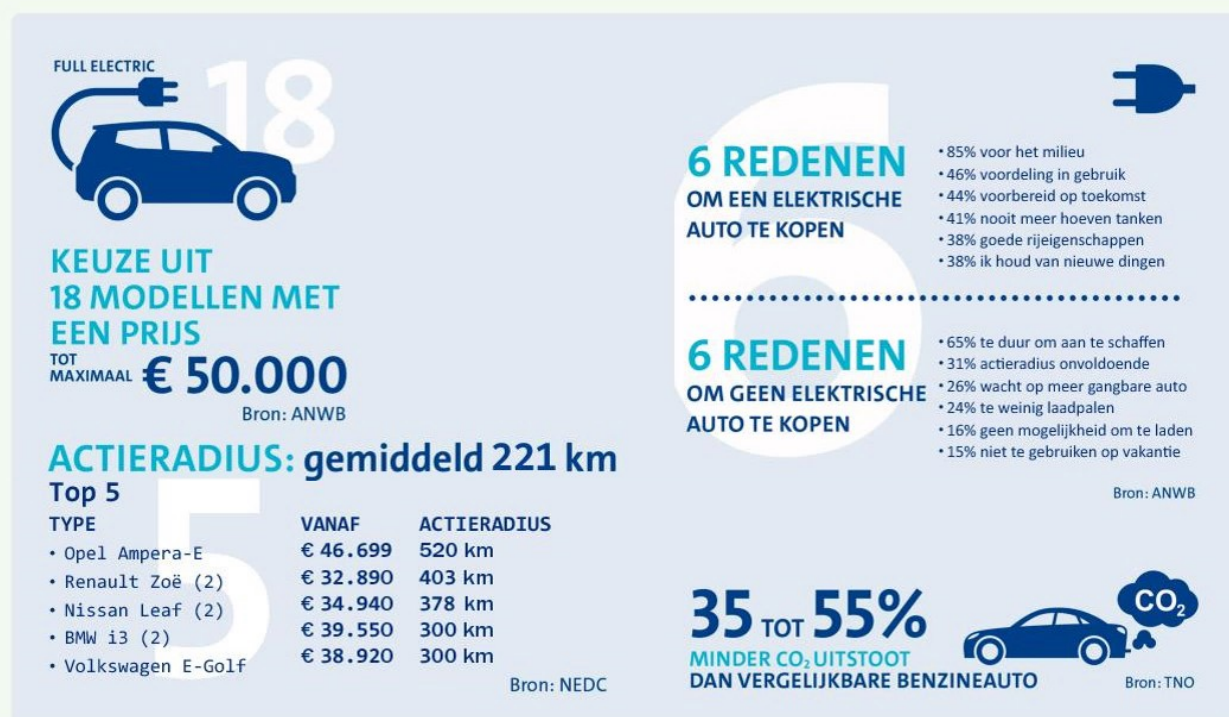


Beginopdracht Elektrisch rijden

De Elektrisch Rijden Monitor van de ANWB geeft jaarlijks de actuele stand van zaken weer rond elektrisch rijden in Nederland. Elektrisch rijden is in opmars, maar komt onder particuliere automobilisten nog steeds weinig voor. De ANWB heeft een representatief onderzoek gedaan onder de Nederlandse bevolking naar de redenen hiervan. Door de uitkomsten van dit onderzoek te koppelen aan data uit de markt krijg je een beeld van de ontwikkeling van elektrisch rijden in Nederland.



Hieronder zie je een deel van de Elektrisch Rijden Monitor 2018.



- De zes redenen om een elektrische auto te kopen en de zes redenen om dat niet te doen, zijn resultaten uit een enquête van de ANWB. Je ziet dat de percentages optellen tot meer dan 100%. Bedenk welke vragen de ANWB gesteld kan hebben om deze resultaten te krijgen. Geef ook aan in welke vorm deze vragen gesteld kunnen zijn.
- Er lijkt een verband te zijn tussen de actieradius en de vanaf prijs. Voor welke auto uit de top 5 is deze verhouding het gunstigst?
- TNO komt tot de conclusie dat er bij een middenklasse elektrische auto per gereden kilometer 170 gram CO₂ vrijkomt. Hoeveel gram CO₂ stoot een vergelijkbare benzineauto dan maximaal uit?

Voorkennis Centrummaten en frequentietabellen

Theorie A Centrummaten

Je kunt gebruikmaken van het gemiddelde om aan te geven waar het centrum van de gegevens zich bevindt. Het gemiddelde is daarom een voorbeeld van een centrummaat. Andere centrummaten zijn de mediaan en de modus.

Je zou misschien verwachten dat bij het gemiddelde de ene helft van de waarnemingsgetallen onder het gemiddelde ligt en de andere helft erboven. Dat hoeft echter niet zo te zijn. De centrummaat die de waarnemingsgetallen wel in twee even grote groepen verdeelt, is de mediaan.

Hieronder zie je hoe je de mediaan berekent bij een oneven en even aantal waarnemingsgetallen.

MEDIAAN BIJ EEN ONEVEN AANTAL

Bij 9 getallen is de mediaan het 5^e getal.

0, 0, 2, 3, 3, 5, 6, 6, 7

mediaan = 3

MEDIAAN BIJ EEN EVEN AANTAL

Bij 10 getallen is de mediaan het gemiddelde van het 5^e en 6^e getal.

0, 0, 2, 3, 3, 5, 6, 6, 7, 8

mediaan = 4

BEREKENEN VAN GEMIDDELDE, MEDIAAN EN MODUS

gemiddelde	Tel alle waarnemingsgetallen bij elkaar op. Deel dit door het aantal waarnemingsgetallen.
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	Ga na of er één waarneming is met de hoogste frequentie. Is die er, dan is dit de modus. Is die er niet, dan is er geen modus.

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

TELAATKOMERS IN KLAS 4HA

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.



Theorie B Frequentietabellen en staafdiagrammen

In de frequentietabel hiernaast staan de (absolute) frequenties bij de bloedgroepen van de leerlingen in klas 4H1. De totale frequentie is $13 + 12 + 2 + 1 = 28$.

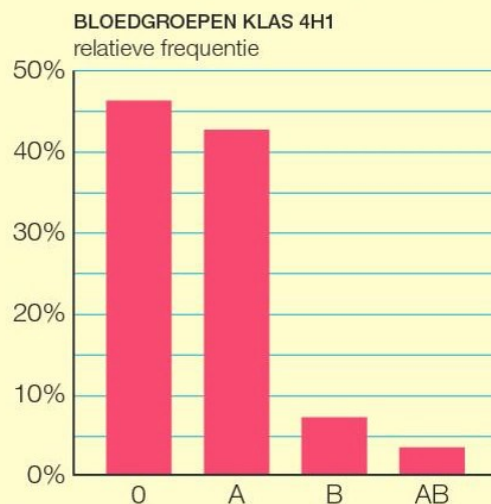
In figuur 2.1 is bij de tabel een staafdiagram getekend. Bij het staafdiagram staan de staven los van elkaar, omdat ze bij verschillende categorieën horen.

In het staafdiagram zijn relatieve frequenties gebruikt. Zo is de relatieve frequentie van bloedgroep 0 gelijk aan

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

BLOEDGROEPEN KLAS 4H1

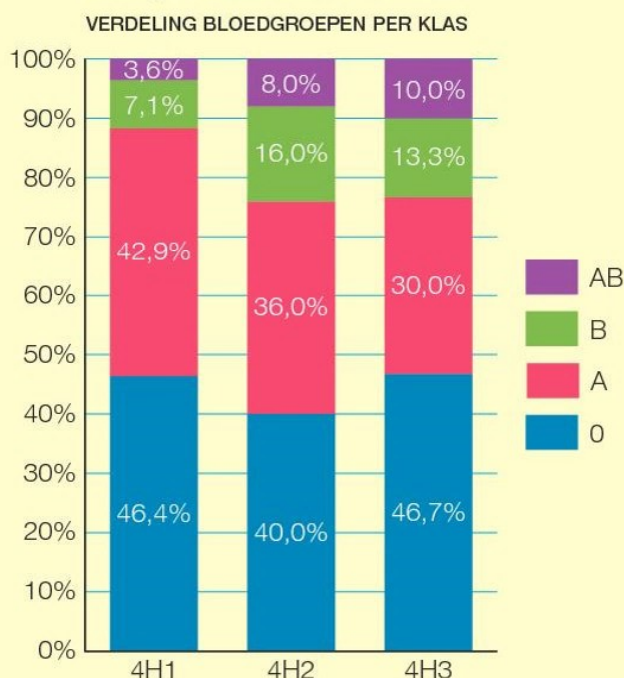
bloedgroep	0	A	B	AB
frequentie	13	12	2	1



figuur 2.1

In figuur 2.2 zie je een samengesteld staafdiagram, ook wel stapeldiagram genoemd.

Je ziet de verdeling van de bloedgroepen in verschillende klassen. Door het gebruik van relatieve frequenties kun je de verdelingen makkelijk vergelijken. Een nadeel is dat je niet meer kunt zien hoeveel leerlingen er in elke klas zitten.



figuur 2.2

2 Zie het samengesteld staafdiagram in figuur 2.2 met de verdeling van de bloedgroepen per klas.

- Welke centrummaat kun je bij de bloedgroepen van klas 4H3 gebruiken? Licht toe.
- In klas 4H2 zitten twee leerlingen met bloedgroep AB. Bereken het aantal leerlingen in 4H2.
- In klas 4H3 zijn er 6 leerlingen meer met bloedgroep A dan met bloedgroep AB. Bereken het aantal leerlingen met bloedgroep 0 in 4H3.

2.1 De onderzoeksvraag



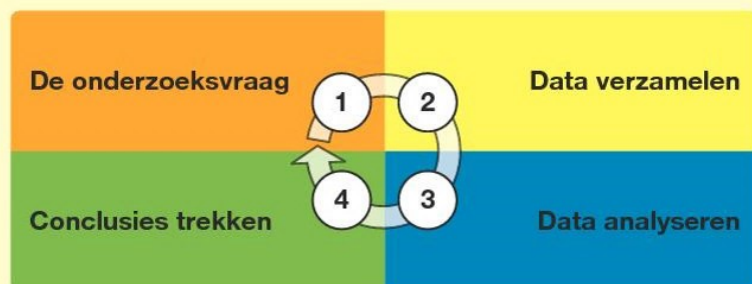
Daan en Emma doen een onderzoek. Hieronder staan zes activiteiten die ze hebben uitgevoerd bij hun onderzoek.

Welke volgorde is logisch?

- 1 Het afvragen of de onderzoeksvraag is beantwoord en zo nodig nieuwe vragen opstellen.
- 2 Het interpreteren van alle verzamelde informatie om te kijken of de onderzoeksvraag hiermee beantwoord kan worden.
- 3 Een overzicht maken van alle aspecten die bij het onderwerp komen kijken, om zo tot de onderzoeksvraag te komen.
- 4 Het beantwoorden van de onderzoeksvraag op basis van de uitkomsten van de verzamelde enquêtes en gevonden informatie.
- 5 Vaststellen welke kenmerken onderzocht moeten worden om de onderzoeksvraag te kunnen beantwoorden.
- 6 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 2.3 De statistische cyclus

In dit hoofdstuk staan de vier fasen van de statistische cyclus centraal. De begrippen en technieken die je in dit hoofdstuk leert zijn allemaal in een of meerdere van deze fasen te gebruiken. Als je zelf onderzoek doet (bijvoorbeeld voor je profielwerkstuk) helpt de statistische cyclus je bij het in goede banen leiden van het proces.



2 Geef bij elk van de stappen uit opgave 1 aan op welke plek ze het beste in de statistische cyclus passen.



3 Lees de onderstaande tekst over kijkonderzoek.

ARTIKEL

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 mensen willen overstappen van een dieselauto naar een elektrische auto?
- b Waarom willen niet meer steden de dieselauto weren uit de binnenstad en waar ligt dat aan?
- c Waarom zijn eigenaren van benzineauto's minder milieubewust dan eigenaren van elektrische auto's?

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 'Wat is het aandeel van de CO₂-uitstoot van het verkeer op de totale CO₂-uitstoot?'.

Het verband kan bovendien *causaal* zijn. Je hebt met een **causaal verband** ofwel **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 de uitstoot van fijnstof door auto's de volksgezondheid in steden?'.

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 mensen die nu in een dieselauto rijden willen binnen drie jaar overstappen op een elektrische auto?’
- 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 ‘Welk verschil in beleid is er tussen Arnhem en Enschede wat betreft het weren van dieselauto’s uit de binnenstad?’.
- De onderzoeksvraag moet *nieuw* zijn.
Het heeft geen zin een onderzoek te doen naar iets dat algemeen bekend is. Een vraag als ‘Hebben elektrische auto’s minder CO₂-uitstoot dan benzineauto’s?’ 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 eigenaren van benzineauto’s minder milieubewust zijn dan eigenaren van elektrische auto’s.
Een betere vraag zou zijn ‘Wat is het verschil in opvattingen over het milieu tussen eigenaren van benzineauto’s en eigenaren van elektrische auto’s?’.

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 ‘Welk verschil in beleid is er tussen Arnhem en Enschede wat betreft het weren van dieselauto’s uit de binnenstad?’ de volgende deelvragen geformuleerd kunnen worden.

- 1** Wat is het beleid van Arnhem wat betreft het weren van dieselauto’s uit de binnenstad?
- 2** Wat is het beleid van Enschede wat betreft het weren van dieselauto’s uit de binnenstad?

Bedenk bij elk van de onderzoeksvragen hieronder twee deelvragen.

- a** Wat is het aandeel van de CO₂-uitstoot van het verkeer op de totale CO₂-uitstoot?
- b** Hoe beïnvloedt de uitstoot van fijnstof door auto’s de volksgezondheid in steden?

Zie opgave 6.

- c** Bedenk bij de onderzoeksvraag die je voor Karlijn hebt 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 beloftes 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.

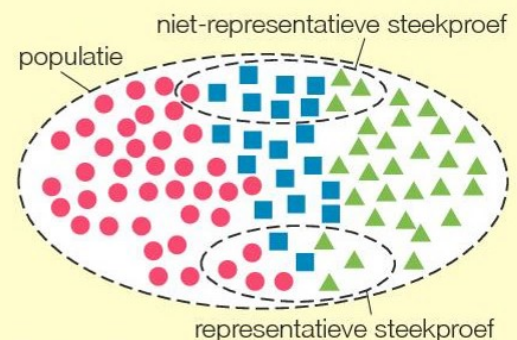
Over de groep Nederlanders in het onderzoek schrijft Stichting KijkOnderzoek (SKO) het volgende.

‘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**. Het onderzoeken van de hele populatie lukt in veel gevallen niet, bijvoorbeeld omdat het teveel werk is. Het deel van de populatie dat je daadwerkelijk onderzoekt heet de **steekproef**. Het is belangrijk dat een steekproef een goede afspiegeling is van de totale populatie. Een steekproef heet dan **representatief**.



figuur 2.4

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 **aselekt** 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 vervolgonopleidingen 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
☐◎*

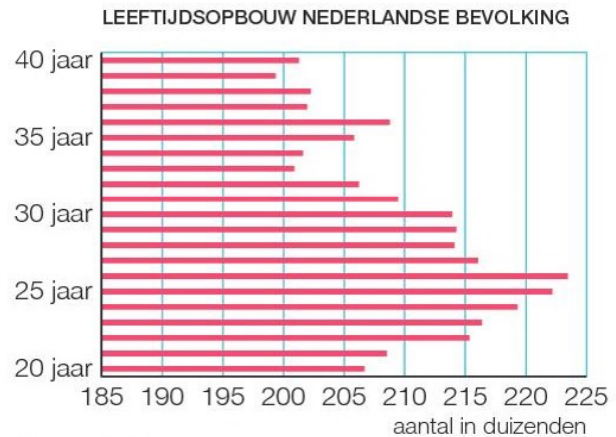
a Er wordt een onderzoek gedaan naar de fruitconsumptie van Nederlanders van 20 tot en met 40 jaar door van elke leeftijd evenveel personen te ondervragen.

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.

014
☐◎*

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 D Populatieproportie en steekproefproportie

Bij de Tweede Kamerverkiezingen worden 150 zetels verdeeld met als uitgangspunt het deel van de stemmen dat elk van de partijen krijgt. De wiskundige term voor deel is **proportie**. Alle stemgerechtigden in Nederland vormen samen de populatie. Wanneer 12% van de populatie op de partij D66 stemt, zeggen we dat de **populatieproportie** D66-stemmers 0,12 is. Deze populatieproportie noteren we als $p = 0,12$. Bij een peiling voor de verkiezingen wordt maar een deel van de stemgerechtigden onderzocht en heb je dus te maken met een steekproef. Wanneer bij een steekproef het aandeel D66-stemmers 9% is, zeggen we dat de **steekproefproportie** 0,09 is. Deze steekproefproportie noteren we als $\hat{p} = 0,09$. Bij een representatieve steekproef zal $\hat{p}$ in de meeste gevallen weinig verschillen van 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

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

- 15** 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.

- A16** 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.

- E17** 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

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 ‘Wat is het aandeel van de CO₂-uitstoot van het verkeer op de totale CO₂-uitstoot?’ is er sprake van een samenhang tussen de CO₂-uitstoot van het verkeer en de totale CO₂-uitstoot.

Bij de vraag ‘Hoe beïnvloedt de uitstoot van fijnstof door auto’s de volksgezondheid in steden?’ is er sprake van een causaal (oorzakelijk) verband: het directe gevolg van de uitstoot van fijnstof door auto’s wordt onderzocht.

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 deel van de populatie dat daadwerkelijk onderzocht wordt heet de steekproef. Om representatief te zijn (een goed beeld van de populatie te geven) moet de steekproef in ieder geval voldoende groot en aselekt zijn.

Bij een aselekte steekproef heeft elk element van de populatie gelijke kans om in de steekproef te komen.

Populatieproportie en steekproefproportie

Wanneer je onderzoekt welk deel van de populatie een bepaalde eigenschap heeft, doe je onderzoek naar een proportie. De proportie is een getal tussen 0 en 1 dat aangeeft welk deel de betreffende eigenschap heeft. Het doel van je onderzoek is het bepalen van de populatieproportie (het deel van de populatie met de eigenschap). Gebruik je een steekproef in je onderzoek, dan geeft dit de steekproefproportie (het deel van de steekproef met de eigenschap). Bij een representatieve steekproef verschillen deze proporties meestal niet veel.

Als 14% van de Nederlanders een zilverkleurige auto heeft, is de proportie Nederlanders met een zilverkleurige auto 0,14.

2.2 Data verzamelen



Hieronder staan twee enquêtevragen.

Leg uit waarom het lastig is deze vragen te beantwoorden.

1 Gebruikt u veel internet?

2 Hoeveel uur per week kijkt u tv?

Kies uit

☐ minder dan 2 uur

☐ meer dan 5 uur

☐ minder dan 5 uur

☐ meer dan 10 uur

Theorie A Enquêtevragen en statistische variabelen

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

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

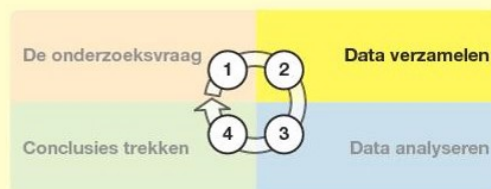
Er zijn verschillende manieren om aan gegevens te komen. Voor veel onderzoeken kan gebruik worden gemaakt van bestaand cijfermateriaal.

Op internet zijn veel gegevens te vinden. Een bekende bron voor gegevens over Nederland en haar inwoners is 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 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.

19
☐◎*

- 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.

020
☐◎*

- 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.

021
☐◎*

Gegeven is de volgende tabel.

BEST BETAALDE SPORTERS 2018

rang	naam	sport	inkomen (\$)	leeftijd
1	Floyd Mayweather	boksen	275 miljoen	41
2	Lionel Messi	voetbal	111 miljoen	31
3	Cristiano Ronaldo	voetbal	108 miljoen	33
4	Conor McGregor	MMA	99 miljoen	30
5	Neymar da Silva Santos Júnior	voetbal	90 miljoen	26

- a Karel zegt ‘Messi en Ronaldo hebben in 2018 samen 219 miljoen dollar verdiend.
Ben je het hier mee eens?’
- b Lars zegt ‘Messi en Ronaldo staan samen op rang 5’.
Ben je het hier mee eens?’
- c Welke kolommen geven een eigenschap weer en welke een hoeveelheid?

Theorie B Statistische variabelen en data

Bij statistiek heten de kenmerken die je onderzoekt **variabelen** omdat ze verschillende waarden kunnen aannemen (ze zijn variabel). De waarden van variabelen zijn niet altijd getallen. Zo kan een variabele bijvoorbeeld de haarkleur zijn met waarden zoals ‘blond’ of ‘bruin’.

Het geheel van alle waarden van de variabelen uit een onderzoek zijn de onderzoeksgegevens oftewel de data.

Bij een variabele zoals haarkleur kun je niet rekenen met verschillen omdat de variabele geen hoeveelheid, maar een eigenschap betreft. Dit soort variabelen heten **kwalitatieve variabelen**.

Variabelen die wel een hoeveelheid betreffen, zoals bijvoorbeeld lichaamslengte in cm, heten **kwantitatieve variabelen**.

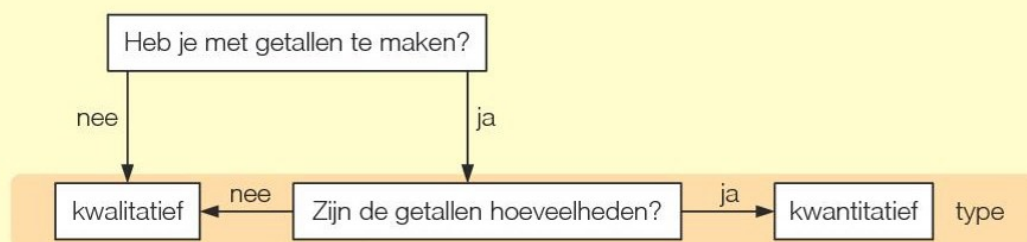
Een ander woord voor eigenschap is *kwaliteit*.
Een ander woord voor hoeveelheid is *kwantiteit*.

Van veel variabelen is direct te zeggen of ze kwalitatief dan wel kwantitatief zijn. Maar soms is het lastiger, zoals bijvoorbeeld bij de variabele rang in opgave 21.

Omdat een rang geen hoeveelheid (kwantiteit) is, kun je bijvoorbeeld niet stellen dat de som van de rangen 1 en 2 gelijk is aan rang 3. De variabele rang betreft een eigenschap en is daarom een kwalitatieve variabele.

De variabele inkomen in opgave 21 betreft een hoeveelheid en is daarom een kwantitatieve variabele.

Bij twijfel kun je als volgt te werk gaan om te bepalen of een variabele kwalitatief of kwantitatief is.



22
□ ⊗ *

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

- | | |
|----------------------|-------------------------|
| I geslacht | III lidmaatschapsnummer |
| II leeftijd in jaren | IV toetscijfer wiskunde |

b Geef nog twee voorbeelden van een kwalitatieve variabele die uitgedrukt is in een getal.

23
□ ⊗ *

Bij een medisch onderzoek worden vaak veel factoren onderzocht. In de tabel hiernaast zie je een deel van een ingevulde vragenlijst. Welke variabelen zijn kwantitatief en welke kwalitatief?

patiëntennummer	11235813
geboortejahr	1985
opleiding	hbo
beroep	leraar
bloedgroep	0
sportfrequentie:	
niet wekelijks	✓
1 – 2 per week	
> 2 keer per week	

R24


Bij een enquête over het aantal dagen per week dat Nederlanders vlees eten, kun je kiezen uit 0 dagen, 1 dag, 2 dagen, 3 dagen en 4 of meer dagen.

De onderzoeker wil weten hoeveel dagen Nederlanders gemiddeld per week vlees eten.

Tegen welk probleem loopt de onderzoeker aan met deze vraag?

O25


Een enquête voor autobezitters bevat de volgende vragen.

I Hoe vaak maakt u gebruik van de auto? Kies uit

☐ nooit ☐ soms ☐ regelmatig ☐ vaak

II Wat voor type auto rijdt u? Kies uit

☐ benzine ☐ diesel ☐ hybride ☐ elektrisch ☐ anders

a Is de variabele autogebruik kwalitatief of kwantitatief? En de variabele type auto?

b Bij welke vraag geven de antwoorden een duidelijke volgorde aan? Licht toe.

Theorie C Meetniveau van een kwalitatieve variabele

Bij kwalitatieve variabelen is het van belang hoe ze zijn gemeten.

Als de waarden een duidelijke volgorde (ordering) hebben, is de variabele gemeten op **ordinaal** niveau. Dat is het geval bij de variabele autogebruik in opgave 25.

Als de volgorde niet van belang is, zoals bij de variabele type auto, is er gemeten op **nominaal** niveau.

De **meetniveaus** nominaal en ordinaal horen bij een kwalitatieve variabele.

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

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

Nominaal is afgeleid van het Latijnse woord *nomen*, dat naam betekent.

MEETNIVEAUS KWALITATIEVE VARIABELE

meetniveau	omschrijving
nominaal	De verschillende categorieën worden <i>benoemd</i> , maar er is geen duidelijke volgorde in de waarden van de variabele.
ordinaal	Er is een duidelijke <i>ordering</i> in de waarden van de variabele.

26


Geef voor de vragenlijst hiernaast bij elke variabele het meetniveau.

Vragenlijst *Omcirkelen wat van toepassing is*

geslacht: man / vrouw
aantal broers/zussen: 0 / 1 / 2 / 3+
vooropleiding: vmbo / havo / vwo
woonplaats:



27
☐⊙

Er wordt onderzoek gedaan naar de vakantiebestemming van Nederlanders in de herfstvakantie.

Bedenk in de context van dit onderzoek een kwantitatieve variabele, een nominale variabele en een ordinale variabele.

028
☐⊙*

Bij een snelheidscontrole heeft de politie te maken met de variabelen 'aantal auto's' en 'snelheid in km/uur'.

Deze variabelen zijn beide kwantitatief, maar toch is er verschil in waarden die deze variabelen kunnen aannemen. Zo kan de variabele 'aantal auto's' bijvoorbeeld niet de waarde 85,6 aannemen.

Omschrijf wat het verschil is in waarden die de variabelen 'aantal auto's' en 'snelheid in km/uur' kunnen aannemen.

Theorie D Discrete en continue variabelen

Een gebouw heeft 12 verdiepingen. Je kunt bijvoorbeeld spreken over de 3^e verdieping of de 4^e verdieping. Maar verdieping $3\frac{1}{2}$ bestaat niet. De verdieping van een gebouw is een voorbeeld van een **discrete variabele**.

De hoogte waarop de lift zich in het gebouw bevindt, kan bijvoorbeeld 3 meter of 4 meter zijn, maar ook elke tussenliggende waarde is mogelijk. De hoogte van de lift in meters is daarom een voorbeeld van een **continue variabele**.

Bij kwantitatieve variabelen maken we onderscheid in discrete en continue variabelen.

- Bij een discrete variabele zijn sommige tussenliggende waarden niet mogelijk.
- Bij een continue variabele zijn alle tussenliggende waarden mogelijk.

29
☐⊙

Geef aan of de variabele continu of discreet is.

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

30
*

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.

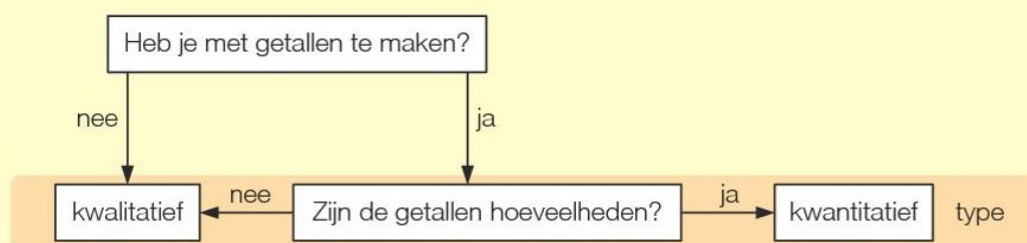
Terugblik

Enquêtevragen en statistische variabelen

Informatie verzamelen kun je op verschillende manieren aanpakken. Wanneer gegevens beschikbaar zijn via bronnen (zoals bijvoorbeeld het CBS) is dat een logisch uitgangspunt. Zijn er geen bronnen dan zul je de gegevens zelf moeten verzamelen, dat kan bijvoorbeeld met een enquête.

Statistische variabelen en data

De kenmerken bij een statistisch onderzoek heten variabelen. Tijdens het onderzoek verzamel je de waarden die de variabelen aannemen. Al deze waarden samen noemen we de onderzoeksgegevens of data. Het is belangrijk de opdeling in kwalitatieve en kwantitatieve variabelen te kennen. Het bepalen van het type variabele doe je met behulp van de mogelijke waarden van de variabele en het schema hieronder.



Nominale en ordinale variabelen

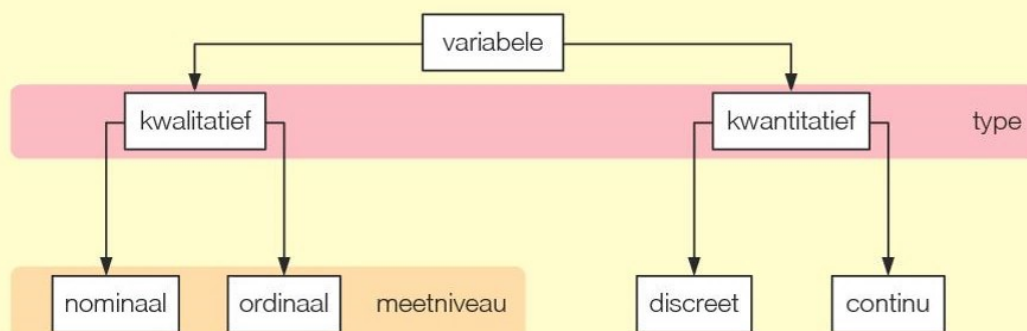
Bij kwalitatieve variabelen is er een onderscheid in de meetniveaus nominaal en ordinaal.

- Bij een ordinale variabele is er een duidelijke ordening in de waarden van de variabele.
- Bij een nominale variabele is er geen duidelijke ordening.

Discrete en continue variabelen

Bij kwantitatieve variabelen maken we onderscheid tussen discrete en continue variabelen.

- Bij een discrete variabele zijn sommige tussenliggende waarden niet mogelijk.
- Bij een continue variabele zijn alle tussenliggende waarden mogelijk.



2.3 Centrum- en spreidingsmaten

031
☐◎*

Bij het invullen van een vragenlijst heb je bij elke vraag de keuze uit de scores hiernaast.

a Van welk meetniveau is hier sprake?

Een onderzoeker laat de vragenlijst door 100 mensen invullen.

b Licht aan de hand van een voorbeeld toe dat het onjuist is als hij de gemiddelde score per vraag berekent.

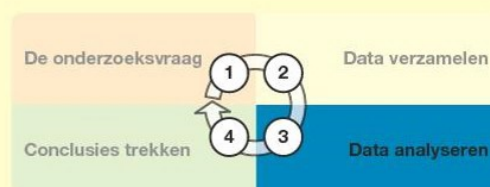
c Welke centrummaat zou hij hier wel kunnen gebruiken?

1:	helemaal mee oneens
2:	mee oneens
3:	acceptabel
4:	mee eens
5:	helemaal mee eens

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 bij een serie waarnemingsgetallen een centrummaat te berekenen.



Verschillende situaties vragen om verschillende centrummaten.

Bij kwalitatieve variabelen kun je de modus gebruiken.

Het voordeel van kwantitatieve variabelen is dat je de keuze hebt uit alle centrummaten.

Let op dat het gemiddelde niet altijd een juiste indruk geeft van het centrum van de waarnemingsgetallen. Het gemiddelde is bijvoorbeeld gevoelig voor uitschieters. Een **uitschieter** is een waarnemingsgetal dat erg afwijkt van de andere waarnemingsgetallen.

VOOR- EN NADELEN BIJ DRIE CENTRUMMATEN

	voordeel	nadeel
gemiddelde	gebruikt alle waarnemingsgetallen	gevoelig voor uitschieters
mediaan	ongevoelig voor uitschieters	alleen de grootte van de middelste waarneming(en) is van belang
modus	geschikt voor elk type variabele	is er niet altijd

32
□ ◎ *

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

- 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.



33
□ ◎

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

34
□ ◎ *

In klas 4HB is nauwgezet het aantal telaarcomers geregistreerd.

TELAATKOMERS IN 4HB

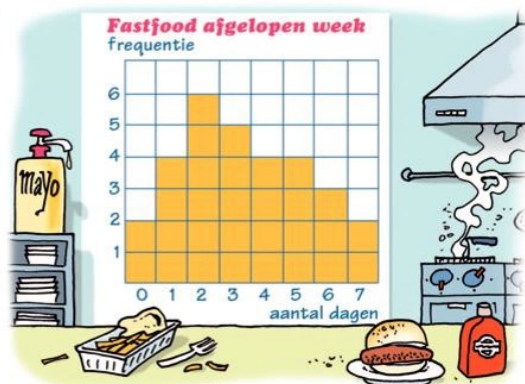
aantal	0	1	2	3	4	5	32
frequentie	1	5	5	3	3	2	1

- a Bereken het gemiddelde.
- b Wanneer de afdelingscoördinator het gemiddelde ziet, is hij erg ontstemd. Hij vindt het gemiddelde aantal telaarcomers onaanvaardbaar hoog. Door het gebruiken van een andere centrummaat weet de klas hem milder te stemmen. Hoe doet de klas dat?

35
□ ⊙ *

Aan een groep leerlingen is gevraagd op hoeveel dagen ze de afgelopen week fastfood hebben gegeten. De eerste 30 resultaten zie je in de figuur hiernaast.

- Met welke variabele heb je hier te maken? Waarom kun je bij deze variabele het gemiddelde, de modus en de mediaan berekenen?
- Zet de centrummaten op volgorde van klein naar groot.
- Op het einde van het onderzoek waren in totaal 40 leerlingen ondervraagd. De mediaan bleek gelijk te zijn gebleven. Het gemiddelde was 3,5 en de modus was groter dan het gemiddelde. Geef een mogelijke frequentietabel bij deze situatie.



figuur 2.7

E36
*

Jori heeft vijf keer een bordspel gespeeld. Zijn scores waren achtereenvolgens 51, 57, 51, 51 en 55.

De maximale score bij dit spel is 60 en de minimale score 50.

Hoe vaak moet hij het spel nog minstens spelen om

- een gemiddelde score van 55 te krijgen
- de modus van zijn scores groter dan het gemiddelde te laten worden
- de mediaan van zijn scores groter dan het gemiddelde te laten worden?

O37
□ ⊙ *

Een docent heeft berekend dat voor een toets de helft van de leerlingen een voldoende heeft behaald.

- Welke centrummaat heeft de docent berekend? Licht toe.

De docent berekent dat de mediaan van de onvoldoendes een 3,6 is en de mediaan van de voldoende een 8,1.

- Hoeveel procent van de leerlingen heeft een cijfer hoger dan een 8,1? En hoeveel procent heeft een cijfer tussen 3,6 en 8,1?

Theorie B De boxplot

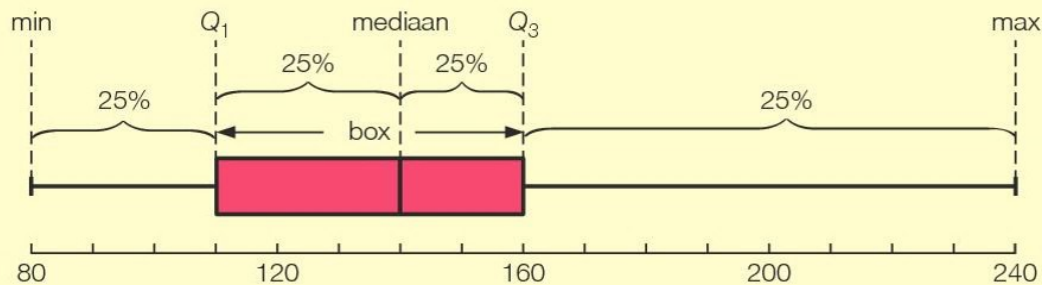
Een **boxplot** is een diagram dat de verdeling van data in beeld brengt.

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 2.8

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.

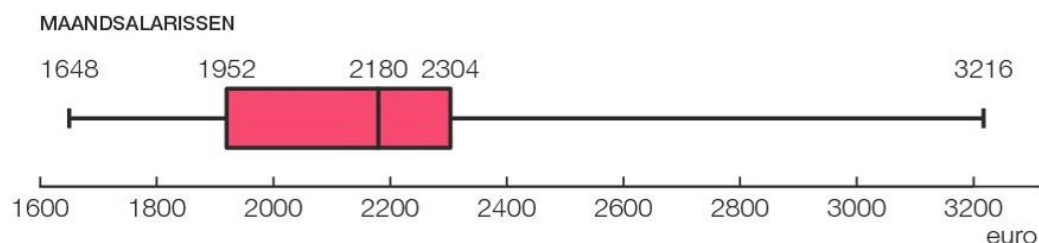
Het verschil tussen het grootste en kleinste waarnemingsgetal heet de **spreadingsbreedte**. In de figuur is de spreidingsbreedte $240 - 80 = 160$. De breedte van de **box** in de boxplot heet de **interkwartielafstand** (ook wel kwartielafstand). In de figuur is de interkwartielafstand $160 - 110 = 50$.

spreadingsbreedte = maximum – minimum

interkwartielafstand = $Q_3 - Q_1$

38

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



figuur 2.9

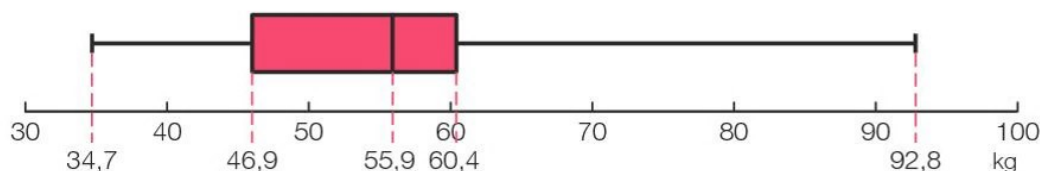
- Bereken de spreidingsbreedte en de interkwartielafstand.
- Hoeveel werknemers hebben een maandsalaris tussen 1952 en 2304 euro?
- Wat weet je van de maandsalarissen van de 25% werknemers met de hoogste maandsalarissen?

39



Van een grote scholengemeenschap zijn de gewichten van alle 2184 leerlingen genoteerd. Er is bij deze data een boxplot gemaakt. Zie figuur 2.10.

GEWICHT VAN LEERLINGEN



figuur 2.10

Een leerling trekt uit bovenstaande gegevens de volgende conclusies. Geef bij elke conclusie aan of deze juist is of onjuist. Licht toe.

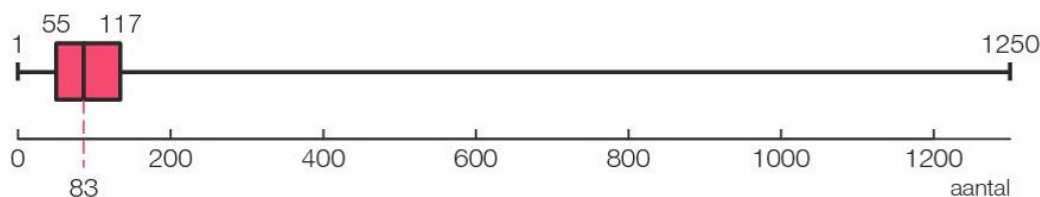
- Er zijn meer leerlingen die een gewicht hebben tussen 60,4 en 92,8 kg dan tussen 55,9 en 60,4 kg.
- Er zijn meer dan 550 leerlingen die lichter zijn dan 46,9 kg.
- Er vallen minder leerlingen binnen dan buiten de box.
- Meer dan driekwart van de leerlingen is lichter dan het gemiddelde van het kleinste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

A40



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

AANTAL KOEIEN PER MELKVEEBEDRIJF IN 2017



figuur 2.11

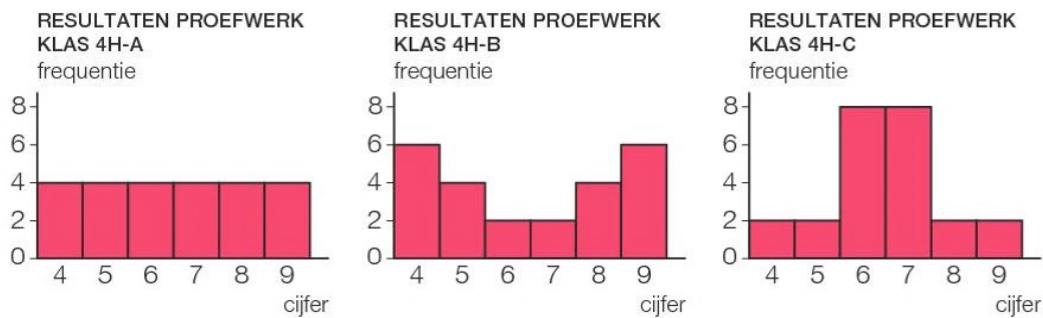
In 2017 waren er 18 056 melkveebedrijven.

- 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 2.12.

Omschrijf dit verschil.



figuur 2.12

Theorie C De standaardafwijking

Twee groepen met gelijke mediaan of gelijk gemiddelde kunnen nog steeds sterk verschillen. Om 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.

42
□ ⊙ *

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 2.13

- 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.

43
□ ⊙ *

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.

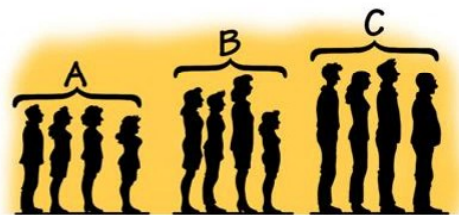
- Hij berekent $\sigma = 5,3$.
Licht toe waarom dit niet kan kloppen.

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.

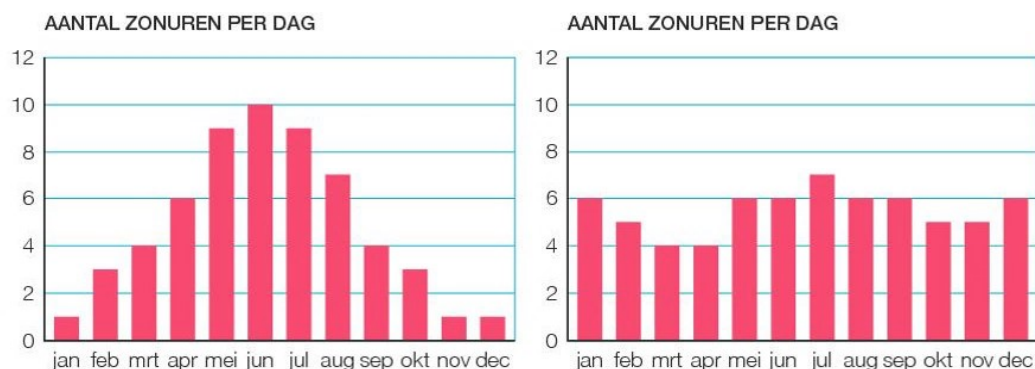
A44
□ ⊙ *

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



figuur 2.14

- 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.



figuur 2.15

Terugblik

Centrummaten

Gemiddelde, mediaan en modus zijn centrummaten.

Elke centrummaat heeft voor- en nadelen. Het hangt van de situatie af welke centrummaat de voorkeur heeft. Zo kun je bijvoorbeeld het gemiddelde niet gebruiken bij een kwalitatieve variabele.

VOOR- EN NADELEN BIJ DRIE CENTRUMMATEN

	voordeel	nadeel
gemiddelde	gebruikt alle waarnemingsgetallen	gevoelig voor uitschieters
mediaan	ongevoelig voor uitschieters	alleen de grootte van de middelste waarneming(en) is van belang
modus	geschikt voor elk type variabele	is er niet altijd

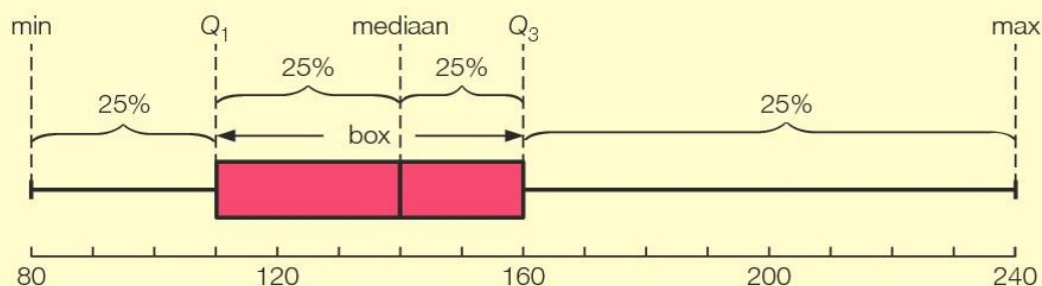
Boxplot

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 .

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 (notatie σ)

De standaardafwijking gebruikt men vaak in combinatie met het gemiddelde.

De interkwartielafstand wordt vaak samen met de mediaan gebruikt.

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

2.4 Representatie van data

O45
☐◎*

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

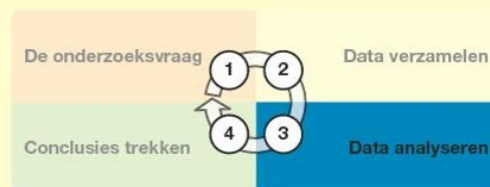
Waarom is het niet zinvol hierbij een frequentietabel te maken?



Theorie A Klassenindeling

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

In opgave 45 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 vastgesteld. Zie de frequentietabel hiernaast.

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

Uitwerking

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

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

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

GEWICHT LEERLINGEN IN KG

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

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

46
☐⊙*

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

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

PRIJS LEVENSMIDDELEN

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

A47
☐

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

- Maak een frequentietabel bij de vloeroppervlakte in m². Neem de klassen $[50, 150)$, $[150, 200)$, ...
- Wat is de modale klasse bij de vloeroppervlakte?
- 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.

A48 Aan de docenten van het Noorderveld College is gevraagd hoe oud ze zijn. Zie de frequentietabel hiernaast.

- a Mevrouw Van Santen is 35 jaar.
In welke leeftijdsklasse valt ze?
- b Een docent Frans zegt dat de gemiddelde leeftijd van de docenten 40 jaar is.
Onderzoek of dit kan kloppen.
- c Kan de mediaan 47 jaar zijn? Licht toe.
- d 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

49 Hiernaast staat het *steel-bladdiagram* bij het aantal werknemers per winkel in een winkelcentrum. Van het aantal 32 op de onderste regel is 3 de steel (het aantal tientallen) en 2 het blad (het aantal eenheden).

- a Geef de modus.
- b Bereken de spreidingsbreedte.
- c Bereken de interkwartielafstand.
- d Een steel-bladdiagram kun je opvatten als een bijzondere klassenindeling.
Wat zijn de klassen van het gegeven steel-bladdiagram?

AANTAL WERKNEMERS	
0	3 4 4 4 6 6 7 7 8 8 8 9 9
1	2 4 6 6 8 8 8 8 9
2	0 2 3 5 8
3	2 3 4
tientallen	eenheden

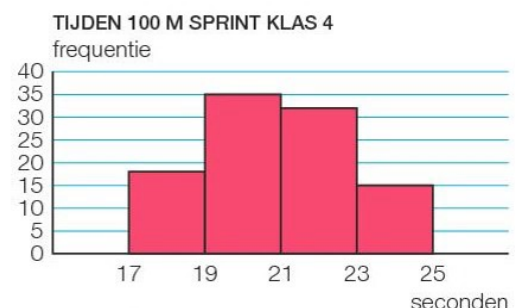
50 Aan een groep ouders is gevraagd op welke leeftijd van de man en de vrouw het eerste kind geboren is. Zie het een *dubbel steel-bladdiagram* hiernaast.

- a Hoeveel ouderparen zijn ondervraagd?
- b Hoeveel mannen van deze groep waren 33 jaar bij de geboorte van het eerste kind? En hoeveel vrouwen waren 35 jaar?
- c Hoeveel mannen waren jonger dan 30 jaar? En hoeveel vrouwen?
- d Luuk maakt een klassenindeling van de mannen en vrouwen samen. Hij neemt als eerste klasse [18, 24>. Hoeveel personen komen er in de klasse [30, 36)? En hoeveel in de hoogste klasse?

LEEFTIJD OUDERS BIJ GEBORTE EERSTE KIND		
man		vrouw
9	1	8 9 9 9
6 5 5 3 3	2	0 0 2 3
6 4 4 3 0	3	5 5
8 4 3	4	0 1 2 2 3 3
1 0	5	
eenheden	tientallen	eenheden

051 Zie het staafdiagram hiernaast.

- a Heb je hier met een kwalitatieve of kwantitatieve variabele te maken?
- b Waarom staan op de horizontale as de getallen bij de grenzen van de staven?
- c Waarom staan de staven niet los van elkaar?

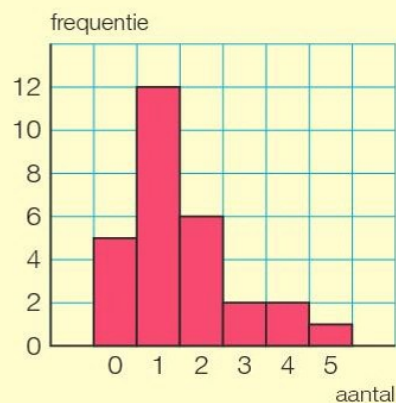


figuur 2.16

Theorie B Histogram en frequentiepolygoon

In de figuur hiernaast is een **histogram** getekend. Bij een histogram staat op de horizontale as een kwantitatieve variabele. Is deze variabele discreet en is er geen klassenindeling gebruikt, dan staan de getallen 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 staafdiagram met een kwantitatieve variabele op de horizontale as. De staven staan tegen elkaar aan.

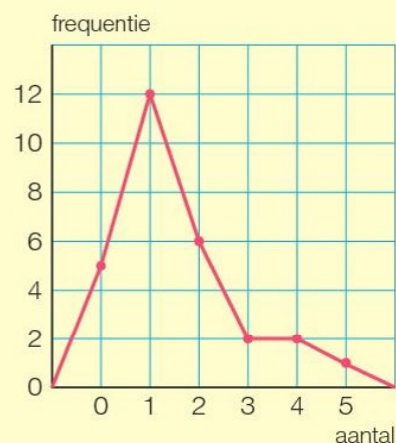


figuur 2.17

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

In figuur 2.18 is bij het histogram in figuur 2.17 de **frequentiepolygoon** getekend. Een frequentiepolygoon is een lijndiagram waarin de frequenties zijn uitgezet tegen de waarnemingsgetallen.

Omdat het hier een discrete variabele betreft waarbij 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. Een frequentiepolygoon begint en eindigt op de horizontale as.



figuur 2.18

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

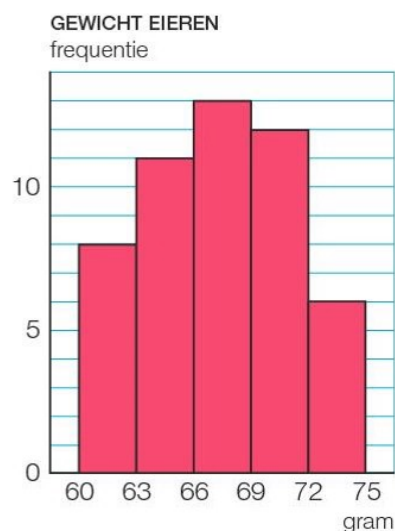
Een frequentiepolygoon is alleen te gebruiken bij een kwantitatieve variabele.

In figuur 2.18 is een frequentiepolygoon getekend bij absolute frequenties. Je kunt dat ook doen bij relatieve frequenties. Je krijgt dan de **relatieve frequentiepolygoon**.

52


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 2.20

53
 *

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

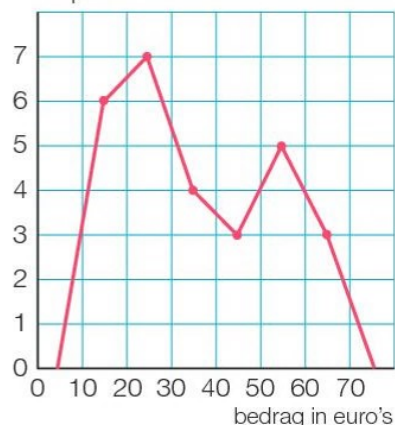
klasse	gewicht in gram	aantal
S	$\langle \leftarrow, 53 \rangle$	441
M	$[53, 63)$	541
L	$[63, 73)$	533
XL	$[73, \rightarrow)$	216

54
 *

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 $[50, 60)$.

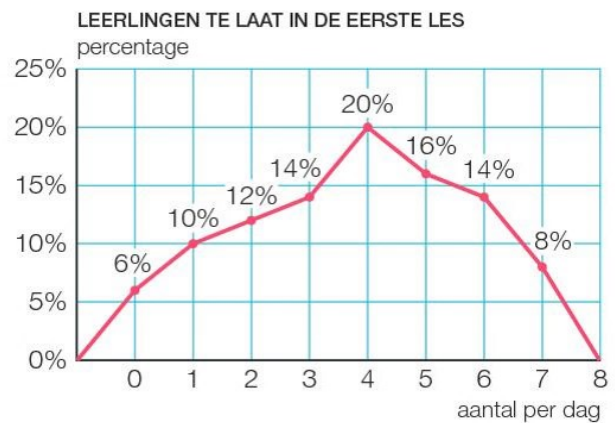
GEMIDDELD BONBEDRAG
frequentie



figuur 2.19

A55
  *

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 2.21

- 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.

O56
  *

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 krijg je de **cumulatieve frequentie**. Wanneer je met cumulatieve frequenties in procenten te maken hebt, noem je het resultaat de **relatieve cumulatieve frequentie**.

Bij 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

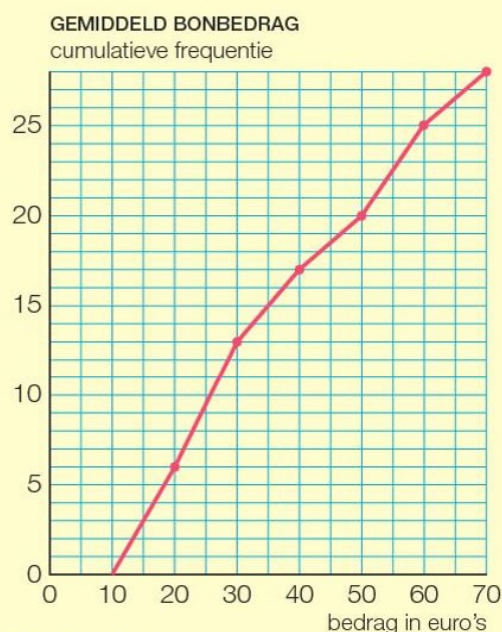
$$\begin{array}{c} 6 \\ \uparrow \\ \text{frequentie} \\ \text{van } [10, 20) \end{array} + \begin{array}{c} 7 \\ \uparrow \\ \text{frequentie} \\ \text{van } [20, 30) \end{array} + \begin{array}{c} 4 \\ \uparrow \\ \text{frequentie} \\ \text{van } [30, 40) \end{array} = 17$$

Er zijn dus 17 winkels waarbij het gemiddeld bonbedrag minder dan 40 euro is. En zo zijn er $6 + 7 + 4 + 3 + 5 + 3 = 28$ winkels waarbij het gemiddeld bonbedrag minder dan 70 euro is.

In figuur 2.22 is de **cumulatieve frequentiepolygoon** getekend. De punten zijn uitgezet boven de *rechtergrens* van elke klasse. Uit deze figuur kun je aflezen dat er 20 winkels zijn met een gemiddeld bonbedrag van minder dan 50 euro. En ook kun je aflezen dat er $28 - 17 = 11$ winkels zijn met een gemiddeld bonbedrag van minstens 40 euro.

Een frequentiepolygoon bij relatieve cumulatieve frequenties heet een **relatieve cumulatieve frequentiepolygoon**.

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 2.22

R57



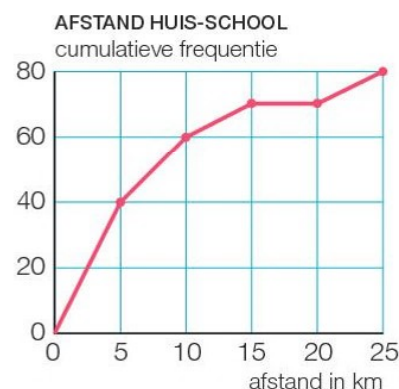
Waarom worden (relatieve) cumulatieve frequenties in een polygoon uitgezet boven de rechterklassengrenzen?

58



Van een groep studenten is de afstand huis-school genoteerd. Zie de cumulatieve frequentiepolygoon in figuur 2.23, 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 20 km van school?
- Hoeveel studenten wonen minstens 10 km, maar minder dan 20 km van school?
- Teken een histogram bij figuur 2.23.

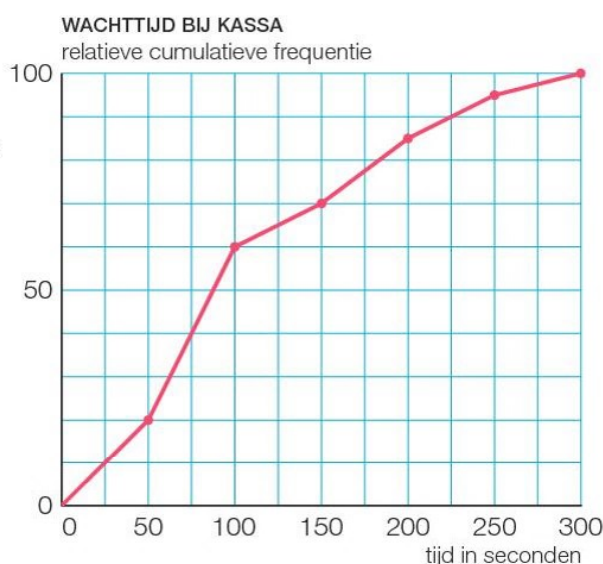


figuur 2.23

59
□ ⊙ *

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 2.24, 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?

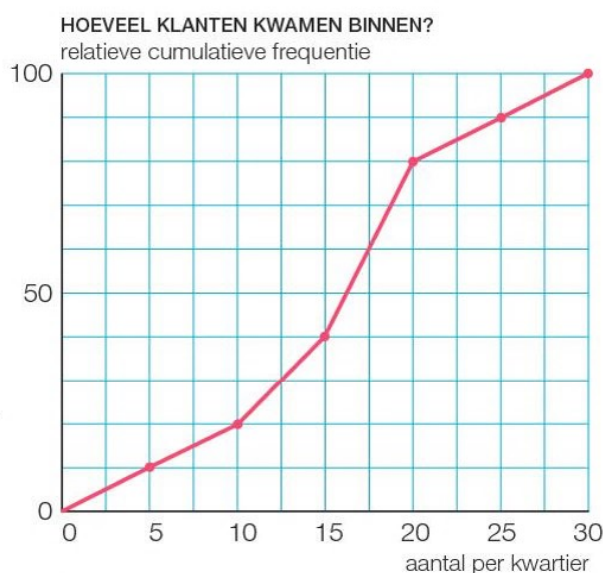


figuur 2.24

A60
□ ⊙ *

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 2.25, 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 2.25 een frequentietabel met absolute frequenties.



figuur 2.25

O61
□ ⊙ *

Het eerste kwartiel, de mediaan en het derde kwartiel verdelen een serie waarnemingsgetallen in vier groepen.

Geef de relatieve cumulatieve frequentie van Q_1 , van de mediaan en van Q_3 .

Theorie D Presentaties in elkaar omzetten

Data van een onderzoek wordt meestal gepresenteerd in diagrammen en tabellen. Daarom wordt een diagram of tabel ook wel een **presentatie** (of **representatie**) van de data genoemd. Welke presentatie de voorkeur heeft hangt af van de data en wat de onderzoeker wil laten zien.

Het is belangrijk presentaties in elkaar om te kunnen zetten zodat met verschillende presentaties gewerkt kan worden. Let op dat na het omzetten soms niet alle oorspronkelijke data meer zichtbaar is. Bij het maken van een boxplot of klassenindeling gaat er bijvoorbeeld informatie verloren. Wanneer je berekeningen maakt op basis van dit soort presentaties krijg je een benadering.

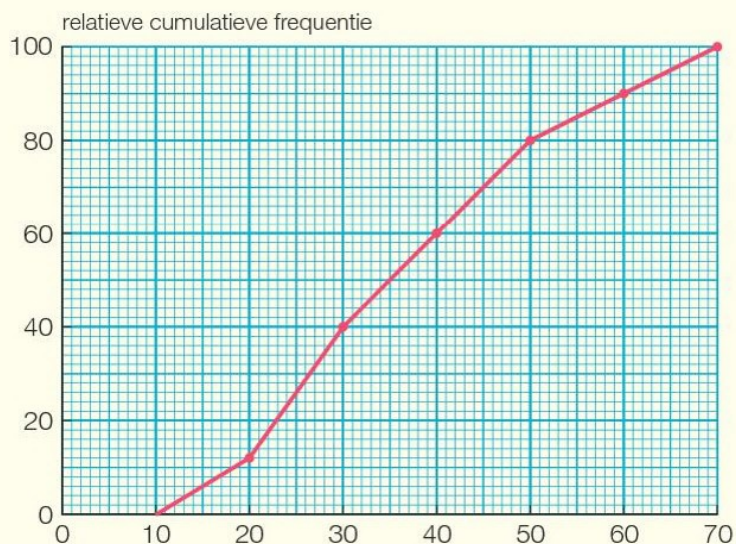
In het voorbeeld wordt bij een relatieve cumulatieve frequentiepolygoon een boxplot gemaakt. Een boxplot geeft een duidelijker overzicht van de spreiding van de gegevens.

Voorbeeld

Zie de relatieve cumulatieve frequentiepolygoon hiernaast. Teken bij deze figuur de boxplot.

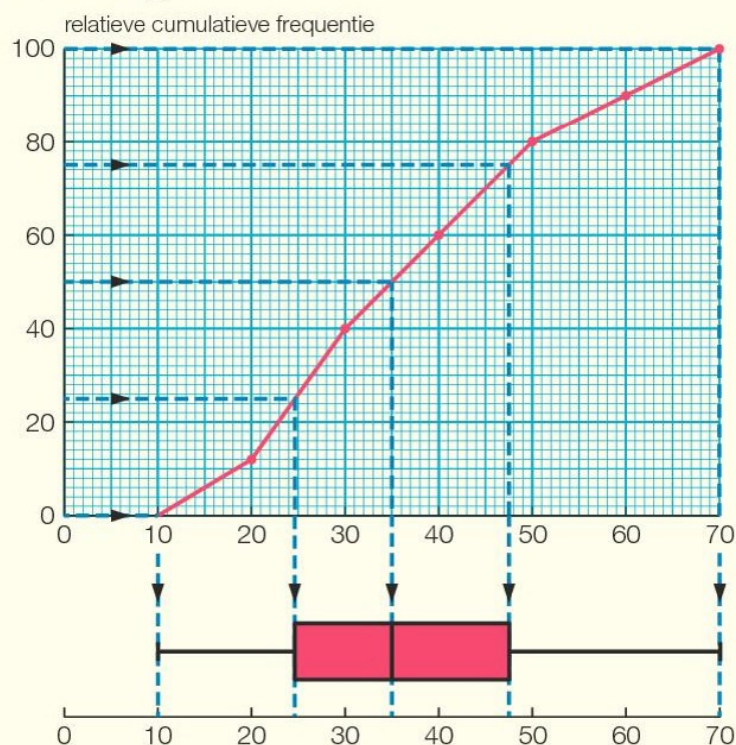
Aanpak

Bepaal de vijf getallen die voor de boxplot nodig zijn, door te kijken welke waarden bij de cumulatieve percentages 0, 25, 50, 75 en 100 horen.



figuur 2.26

Uitwerking



Laat zien hoe je te werk gaat bij het bepalen van de waarden uit een diagram. Hiernaast is dat gedaan met stippellijnen.

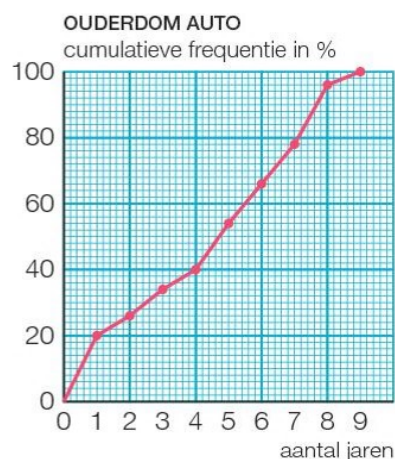


Zie het voorbeeld. In de boxplot is de mediaan bij 35 getekend. De mediaan van de oorspronkelijke data is echter 38. Verklaar dit verschil.



[▶ WERKBLAD] Tijdens een autoshow wordt een groot aantal tweedehands auto's aangeboden. Figuur 2.27 geeft informatie over de ouderdom van de aangeboden auto's.

- Bepaal met behulp van figuur 2.27 de waarden van Q_1 , de mediaan en Q_3 .
- Teken de boxplot bij figuur 2.27.
- Onderzoek of een kwart van de auto's in de autoshow ouder dan 6 jaar en 2 maanden kan zijn.

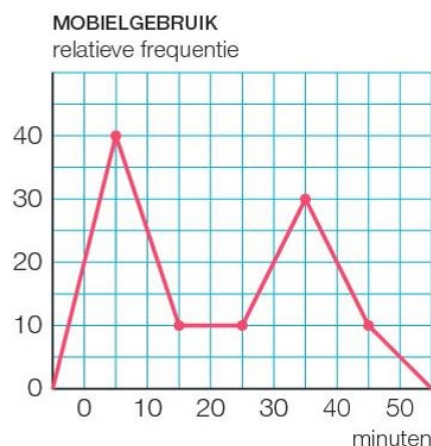


figuur 2.27



Van een aantal leerlingen is bijgehouden hoeveel minuten ze tijdens een pauze hun mobiel gebruikten. De resultaten zijn verwerkt in de relatieve frequentiepolygoon in figuur 2.28, waarbij de eerste klasse $[0, 10)$ is.

- Teken de boxplot bij figuur 2.28.
- Welke interkwartielafstand volgt uit de boxplot?
- Onderzoek of de werkelijke interkwartielafstand 38 minuten kan zijn.



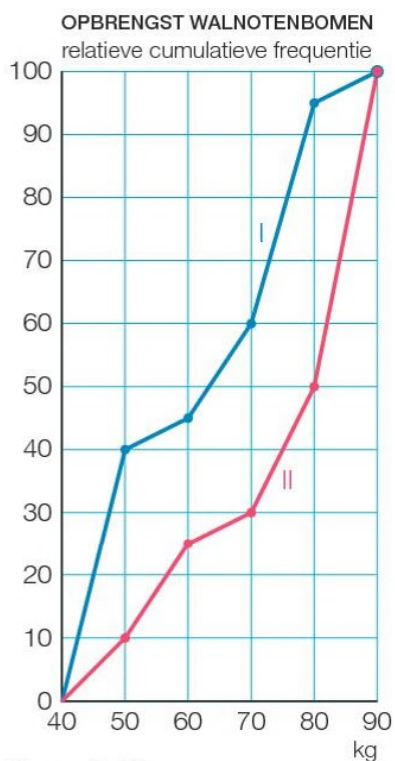
figuur 2.28



Pietro Pallina bezit twee stukken land met walnotenbomen. Op perceel I staan 200 bomen, op perceel II staan 160 bomen.

De opbrengst per boom in kg is verwerkt in figuur 2.29, waarbij de eerste klasse $[40, 50)$ is.

- Teken de relatieve frequentiepolygoon voor perceel I.
- Hoeveel bomen van perceel I brachten minstens 60, maar minder dan 70 kg op?
- Teken het histogram met absolute frequenties voor perceel II.
- Benader met behulp van de klassenmiddens de gemiddelde opbrengst per boom op perceel II.



figuur 2.29

Terugblik

Steel-bladdiagram en klassenindeling

Een manier om (kleine hoeveelheden) data overzichtelijk te ordenen is een steelbladdiagram. Van het getal 40 op de onderste regel is 4 de steel en 0 het blad.

STEEL-BLADDIAGRAM

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

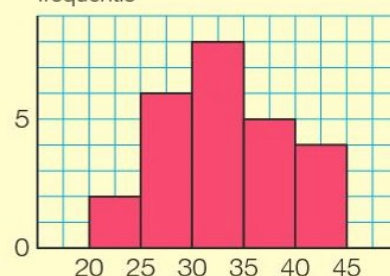
klasse	frequentie
[20, 25)	2
[25, 30)	6
[30, 35)	8
[35, 40)	5
[40, 45)	4

Hiernaast staat een klassenindeling bij deze data.

Bij de klasse $[20, 25)$ doet de linkergrens 20 mee, maar de rechtergrens 25 niet. De klassenbreedte van deze klasse is 5 en het klassenmidden is 22,5. Een nadeel van een klassenindeling is het verlies aan informatie. Zo kun je in de klasse $[20, 25)$ niet meer zien welke waarde de twee metingen hadden.

Maak je bij een klassenindeling een histogram, dan staan de klassengrenzen tussen de staven, zie hiernaast.

HISTOGRAM
frequentie



Frequentiepolygoon en cumulatieve frequentiepolygoon

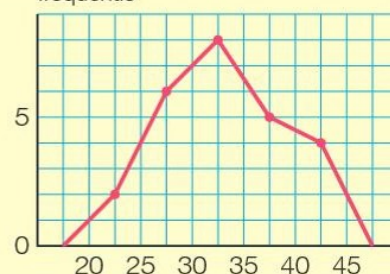
Bij een frequentietabel van een kwantitatieve variabele is een frequentiepolygoon te maken.

Een frequentiepolygoon is een lijndiagram waarin frequenties zijn uitgezet tegen de waarnemingsgetallen. Begin- en eindpunt liggen op de horizontale as. Bij een relatieve frequentiepolygoon zet je de relatieve frequenties uit tegen de waarnemingsgetallen.

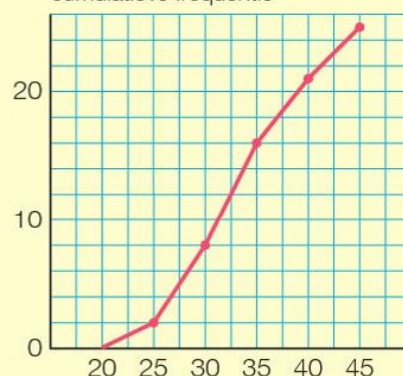
Door bij de frequentie van een waarneming alle frequenties van voorgaande waarnemingen op te tellen, krijg je de cumulatieve frequentie. Bij de klasse $[30, 35)$ hoort de cumulatieve frequentie $2 + 6 + 8 = 16$.

Een frequentiepolygoon bij cumulatieve frequenties heet een cumulatieve frequentiepolygoon. Let erop dat de cumulatieve frequentie altijd boven de *rechterklassengrens* wordt uitgezet.

FREQUENTIEPOLYGOON
frequentie



CUMULATIEVE
FREQUENTIEPOLYGOON
cumulatieve frequentie



Diagrammen omzetten

Diagrammen helpen om data te presenteren. Een diagram wordt daarom ook wel een presentatie of representatie van de data genoemd.

Op deze pagina staat een aantal representaties van dezelfde data. Let op dat bij het maken van een boxplot of een klassenindeling informatie verloren gaat.

2.5 Conclusies trekken

O66




Bij het examenuitje in havo 5 kunnen de leerlingen kiezen tussen zeilen of naar Disneyland Parijs. In de tabel hiernaast zie je wat de leerlingen hebben gekozen.

- Met welke variabelen heb je in de tabel te maken? Zijn deze variabelen kwalitatief of kwantitatief? Licht toe.
- Vind je dat er een groot, middelmatig of gering verschil is tussen de keuze van de jongens en de meisjes?

	bestemming		
	zeilen	Disneyland	
jongen	21	18	39
meisje	15	31	46
	36	49	85

Theorie A Verschillen kwantificeren bij nominale variabelen

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

Je leert conclusies trekken over de grootte van een verschil tussen twee groepen. Om zo'n conclusie te trekken bereken je eerst een **associatiemaat**.

Een associatiemaat drukt in een getal uit hoeveel twee groepen van elkaar verschillen. Een ander woord voor in een getal uitdrukken is **kwantificeren**.

In opgave 66 staat een tabel met de nominale variabele geslacht, die de waarden jongen en meisje heeft, en de nominale variabele bestemming, die de waarden zeilen en Disneyland heeft. Een tabel als deze, die de frequentie weergeeft bij elke combinatie van waarden, heet een **kruistabel**. Vaak worden ook de totale frequenties bij de kruistabel genoteerd.

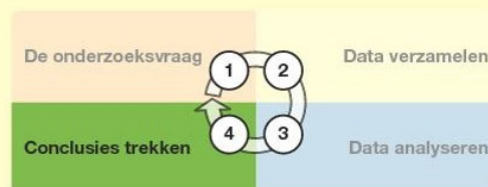
Bij twee nominale variabelen met elk twee waarden gebruiken we als associatiemaat de **phi-coëfficiënt**.

Bij de 2×2 kruistabel hiernaast bereken je de phi-coëfficiënt met

$$\text{phi} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Bij de phi-coëfficiënt hanteren we de volgende vuistregels om een conclusie te trekken over de grootte van het verschil tussen twee groepen.

- Bij $\text{phi} < -0,4$ of $\text{phi} > 0,4$ zeggen we *het verschil is groot*.
- Bij $-0,4 \leq \text{phi} < -0,2$ of $0,2 < \text{phi} \leq 0,4$ zeggen we *het verschil is middelmatig*.
- Bij $-0,2 \leq \text{phi} \leq 0,2$ zeggen we *het verschil is gering*.

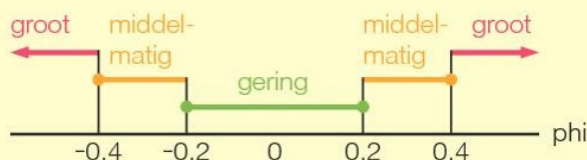


Het woord associatie wordt gebruikt wanneer verschillende dingen met elkaar in verband worden gebracht.

	waarde I	waarde II	
waarde 1	a	b	$a + b$
waarde 2	c	d	$c + d$
	$a + c$	$b + d$	

Hiernaast staan deze vuistregels schematisch weergegeven.

VUISTREGELS PHI-COËFFICIËNT



De associatiematen en vuistregels voor het bepalen van de grootte van het verschil staan bij het centraal examen op een formuleblad. Je hoeft ze dus niet uit het hoofd te leren.

Voorbeeld

Er is onderzoek gedaan naar de tevredenheid over de woonsituatie. Daarbij is onderscheid gemaakt tussen huurwoningen en koopwoningen. Zie de tabel.

Onderzoek met behulp van de phi-coëfficiënt of het verschil in tevredenheid groot, middelmatig of gering is.

	tevredenheid		
	ontevreden	tevreden	
huurwoning	485	3172	3657
koopwoning	180	6163	6343
	665	9335	10 000

Uitwerking

$$\phi = \frac{485 \cdot 6163 - 180 \cdot 3172}{\sqrt{3657 \cdot 6343 \cdot 665 \cdot 9335}} \approx 0,202$$

Let op met afronden, want dat kan de conclusie beïnvloeden.

Omdat $0,2 < \phi \leq 0,4$ is het verschil gering.

67 Bereken bij opgave 66 met behulp van de phi-coëfficiënt of het verschil in bestemming tussen jongens en meisjes groot, middelmatig of gering is.

68 Bij sommige medicijnen hangt het al of niet optreden van bijwerkingen samen met de aanwezigheid van een genetische factor bij de patiënt. In deze opgave bekijken we een onderzoek onder 80 patiënten die een bepaald medicijn gebruiken. Van de patiënten hebben 29 de genetische factor die wordt onderzocht. Van de 32 patiënten die bijwerkingen hadden bleek bij 17 de genetische factor aanwezig te zijn.

- Vul de kruistabel in.
- Onderzoek met de phi-coëfficiënt of het verschil in bijwerkingen tussen de patiënten met en de patiënten zonder de genetische factor groot, middelmatig of gering is.

		bijwerkingen	
		wel	niet
genetische factor	wel		
	niet		

A69
☐◎*

‘Windmolens voor de kust verpesten mijn strandbezoek’. Dat is de stelling die aan 2081 Nederlandse en 512 Duitse toeristen werd voorgelegd. Het Duitse deel van het onderzoek werd gedaan door onderzoeksbureau GfK (Growth from Knowledge) in opdracht van Natuur en Milieu. Het bleek dat 620 Nederlanders het eens waren met de stelling, terwijl dit aantal bij de Duitse toeristen maar 72 was. Onderzoek met behulp van de phi-coëfficiënt of het verschil in mening tussen Duitse en Nederlandse toeristen groot, middelmatig of gering is.



E70
*

Bij een discussie wordt een gezelschap mensen verdeeld in twee even grote groepen. In de ene groep is 75% voor de stelling, terwijl in de andere groep 67% tegen de stelling is. Onderzoek met de phi-coëfficiënt of je hier te maken hebt met een groot, middelmatig of klein verschil.

O71
☐◎*

Fenne heeft onderzoek gedaan naar de relatie tussen werksector en besteedbaar inkomen. Voor de werksector is onderscheid gemaakt tussen werken bij de overheid of in een bedrijf. Het inkomen is ingedeeld in vijf categorieën waarbij 1 het laagste inkomen is en 5 het hoogste. Zie de tabel.

- Wat is het meetniveau van de variabele inkomen?
- Waarom kun je hier niet de phi-coëfficiënt gebruiken om het verschil te kwantificeren?

inkomens- categorie	werksector	
	overheid	bedrijf
1	25	11
2	58	34
3	24	32
4	7	26
5	6	12

Theorie B Verschillen kwantificeren bij ordinale variabelen

Bij een ordinale variabele gebruik je de associatiemaat **maximale verschil in cumulatief percentage** (max.Vcp) om het verschil te kwantificeren.

De max.Vcp bereken je bij de gegevens van opgave 71 als volgt.

- Bereken voor beide werksectoren bij elke inkomenscategorie de cumulatieve frequentie.
- Bereken voor beide werksectoren bij elke inkomenscategorie het cumulatieve percentage, oftewel de relatieve cumulatieve frequentie.
- Bereken bij elke inkomenscategorie het verschil van de cumulatieve percentages.
- Kijk wat het grootste verschil is, dat is de max.Vcp.

In de tabel hieronder is deze berekening uitgevoerd.

	werksector				
	overheid		bedrijf		
inkomens- categorie	cum. freq.	cum. perc.	cum. freq.	cum. perc.	Vcp
1	25	20,8%	11	9,6%	11,2%
2	83	69,2%	45	39,1%	30,1% ← max.Vcp
3	107	89,2%	77	67,0%	22,2%
4	114	95,0%	103	89,6%	5,4%
5	120	100%	115	100%	0%

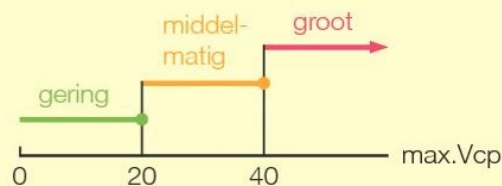
Bij de max.Vcp hanteren we de volgende vuistregels om een conclusie te trekken over de grootte van het verschil tussen twee groepen.

- Bij $\text{max.Vcp} > 40$ zeggen we *het verschil is groot*.
- Bij $20 < \text{max.Vcp} \leq 40$ zeggen we *het verschil is middelmatig*.
- Bij $\text{max.Vcp} \leq 20$ zeggen we *het verschil is gering*.

Hiernaast staan deze vuistregels schematisch weergegeven.

In de tabel zie je dat het maximale verschil in cumulatief percentage 30,1% is. Het verschil in inkomenscategorie tussen overheid en bedrijfsleven is middelmatig, want $20 < \text{max.Vcp} \leq 40$.

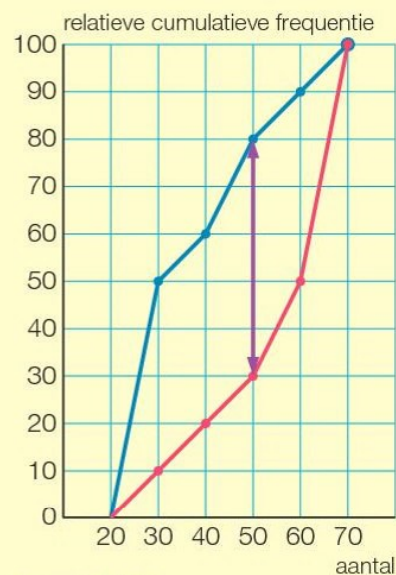
VUISTREGELS MAX.VCP



Zijn de gegevens gepresenteerd in bijvoorbeeld een relatieve cumulatieve frequentiepolygoon of een stapeldiagram, dan kun je de relatieve cumulatieve frequentie direct aflezen. Je hoeft dan dus geen tabel te maken om de max.Vcp te berekenen.

Zo vind je de max.Vcp in figuur 2.30 door te zoeken naar de grootste verticale afstand tussen de twee relatieve cumulatieve frequentiepolygonen. Dit geeft $\text{max.Vcp} = 80 - 30 = 50\%$.

Het verschil is dus groot, want $\text{max.Vcp} > 40$.



figuur 2.30

72
□ ⊗ *

Carglass is een bedrijf dat autoruiten repareert. Na een reparatie wordt de klant gevraagd mee te werken aan een klanttevredenheidsonderzoek. In de tabel hieronder zie je hoe in het vierde kwartaal van 2019 de klanten van de vestigingen in Arnhem en Nijmegen hebben geantwoord bij de stelling “Ik ben snel en deskundig geholpen”.

IK BEN SNEL EN DESKUNDIG GEHOLPEN

	Arnhem	Nijmegen
zeer mee oneens	5	3
mee oneens	6	1
neutraal	12	18
mee eens	38	66
zeer mee eens	64	32

Onderzoek of het verschil tussen de vestigingen in Arnhem en Nijmegen wat betreft het beantwoorden van deze stelling groot, middelmatig of gering is.

73
□ ⊗ *

Voor twee hotels staan de beoordelingen in het gestapelde staafdiagram hiernaast vergeleken.

- Wat is het meetniveau van de variabele beoordeling?
- Onderzoek of het verschil in beoordeling tussen de hotels groot, middelmatig of gering is.



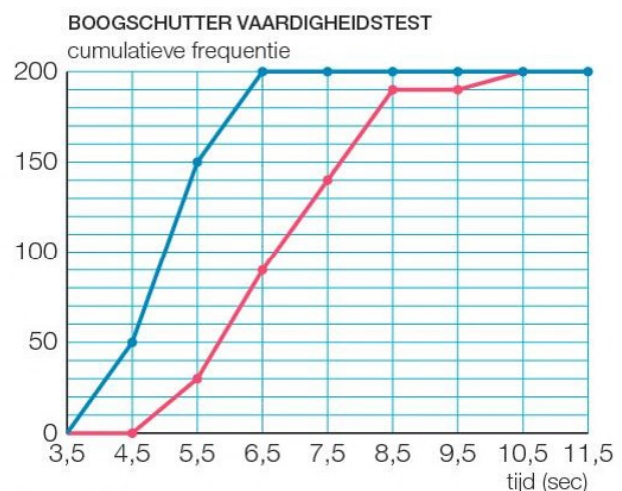
figuur 2.31

A74
□ ⊗ *

Bij een vaardigheidstest wordt een groep boogschutters gevraagd om zo snel mogelijk een doelwit te raken. De proefpersonen zijn verdeeld in twee groepen:

Groep 1 had een klein doelwit.
Groep 2 had een groot doelwit.
De resultaten van het onderzoek zijn verwerkt in de figuur hiernaast.

- Welke cumulatieve frequentiepolygoon hoort bij het kleine doelwit? Licht toe.
- Onderzoek of het verschil in snelheid tussen de groepen groot, middelmatig of gering is.



figuur 2.32

Twee groepen met wiskunde A leerlingen hebben verschillend gescoord voor een toets. Het blijkt dat het eerste kwartiel van de ene groep gelijk is aan het derde kwartiel van de andere groep.

Vind je dat er een groot, middelmatig of gering verschil is tussen de resultaten van de twee groepen?

Theorie C Verschillen bepalen bij kwantitatieve variabelen

Voor het bepalen van de grootte van het verschil bij kwantitatieve variabelen leer je twee manieren. Deze manieren zijn **twee boxplots vergelijken** en de **effectgrootte** berekenen.

Twee boxplots vergelijken

Voor het vergelijken van twee boxplots gebruik je het eerste kwartiel, de mediaan en het derde kwartiel.

Het gedeelte tussen het eerste en het derde kwartiel heet de box.

Bij vergelijken van twee boxplots hanteren we de volgende vuistregels om een conclusie te trekken over de grootte van het verschil tussen twee groepen.

- Als de boxen elkaar niet overlappen, dan zeggen we *het verschil is groot*.
- Als de boxen elkaar wel overlappen en een mediaan van een boxplot ligt buiten de box van de andere boxplot, dan zeggen we *het verschil is middelmatig*.
- In alle andere gevallen zeggen we *het verschil is gering*.

Effectgrootte

Wanneer bij twee groepen zowel het gemiddelde als de standaardafwijking bekend is, gebruik je als associatiemaat de effectgrootte.

Je berekent de effectgrootte met de formule $E = \frac{\bar{X}_1 - \bar{X}_2}{\frac{1}{2}(S_1 + S_2)}$.

$\bar{X}_1$ en $\bar{X}_2$ zijn de gemiddelden van groep 1 en groep 2.

De bijbehorende standaardafwijkingen zijn S_1 en S_2 .

Neem $\bar{X}_1 > \bar{X}_2$, zodat de effectgrootte positief is.

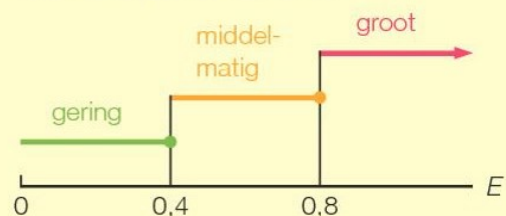
Bij effectgrootte hanteren we de volgende vuistregels om een conclusie te trekken over de grootte van het verschil tussen twee groepen.

- Bij $E > 0,8$ zeggen we *het verschil is groot*.
- Bij $0,4 < E \leq 0,8$ zeggen we *het verschil is middelmatig*.
- Bij $E \leq 0,4$ zeggen we *het verschil is gering*.

Hiernaast staan deze vuistregels schematisch weergegeven.

Benieuwd waarom hier de letter S wordt gebruikt in plaats van σ ? Zie het informatief op de volgende bladzijde.

VUISTREGELS EFFECTGROOTTE



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 beschikbaar is, 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.

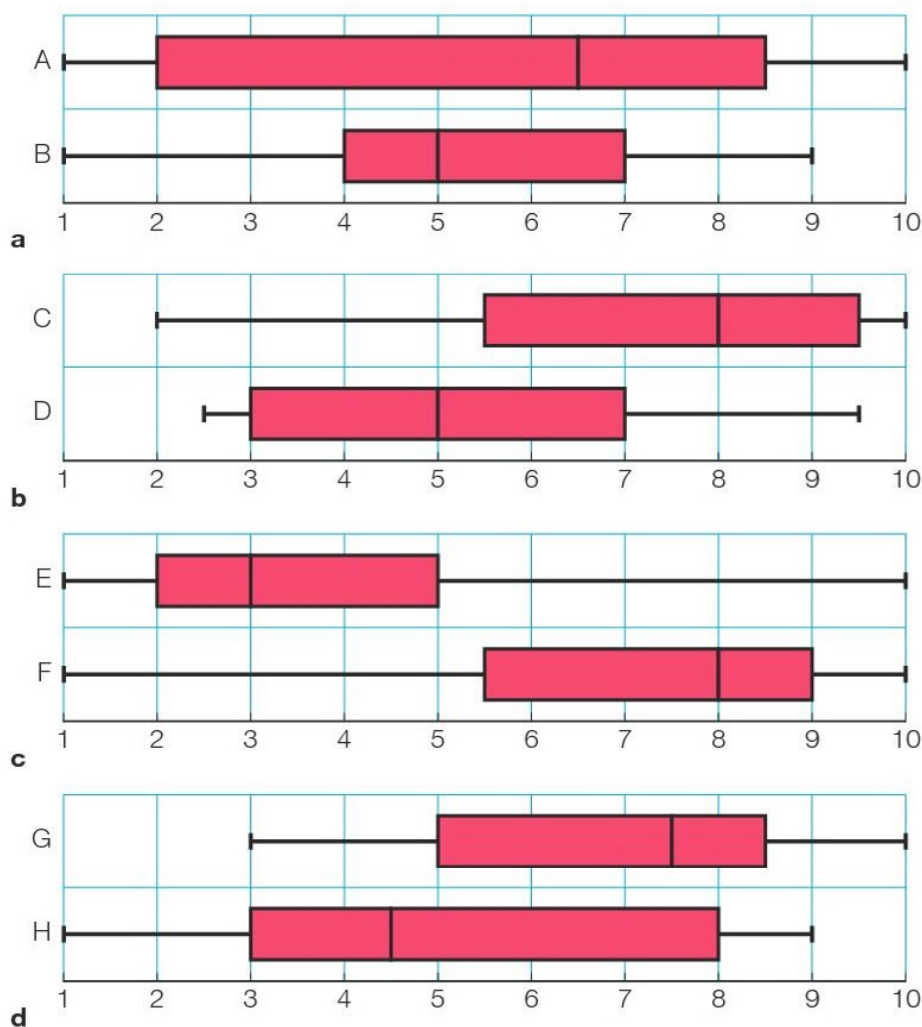
Hoe groter de steekproef, hoe kleiner het verschil is tussen σ en S .

In de opgaven gebruik je de gegeven standaardafwijking.

76
☐⊙*

In de figuur hieronder zie je vier keer de boxplots van twee groepen.

Vergelijk de boxplots en noteer of het verschil tussen de boxplots groot, middelmatig of gering is.



figuur 2.33

Van twee merken appelsap is de inhoud van een aantal pakken vergeleken. De resultaten staan in de tabel.

INHOUD 1 LITERPAK APPELSAP

	gem.	stand. afw.
merk A	1007 mL	3 mL
merk B	1012 mL	10 mL

- Welke variabelen zijn hier onderzocht? Zijn dit kwalitatieve of kwantitatieve variabelen?
- Onderzoek met een associatiemaat of het verschil in inhoud tussen de merken groot, middelmatig of gering is.

[▶WERKBLAD] In de figuur hiernaast zijn nogmaals de gegevens over de opbrengst van walnotenbomen op de twee percelen van Pietro Pallina weergegeven.

- Onderzoek met de max. Vcp of het verschil in opbrengst groot, middelmatig of gering is.

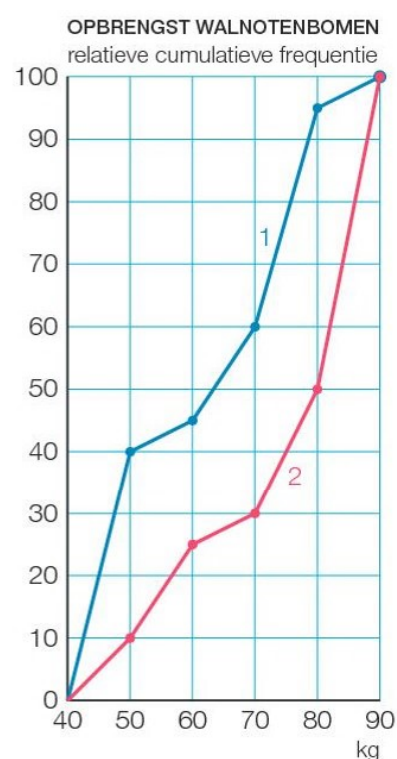
Het is ook mogelijk een conclusie te trekken over het verschil in opbrengst door twee boxplots te vergelijken.

- Teken de boxplots en gebruik deze om te onderzoeken of het verschil in opbrengst groot, middelmatig of gering is.

Een onderzoeker heeft op basis van de oorspronkelijke data berekend dat $\bar{X}_1 = 62,0$ en $S_1 = 14,4$ en dat $\bar{X}_2 = 73,5$ en $S_2 = 14,2$.

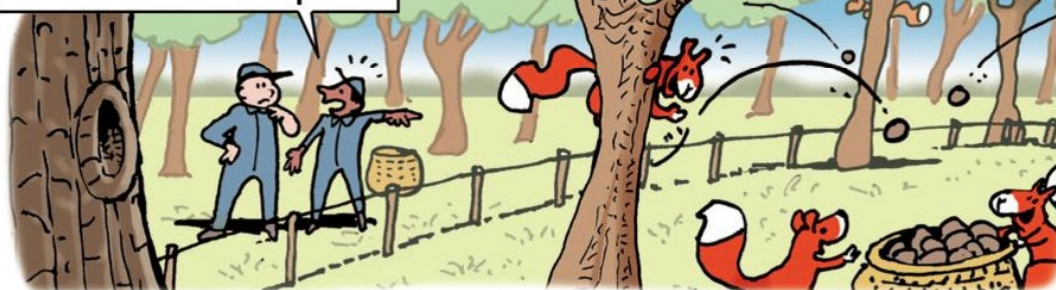
- Onderzoek met de effectgrootte of het verschil in opbrengst groot, middelmatig of gering is.
- Welke van de drie conclusies over het verschil in opbrengst is het meest betrouwbaar? En welke het minst betrouwbaar?

Licht je antwoorden toe.



figuur 2.34

Meneer Pallina, dat andere perceel brengt veel minder op.



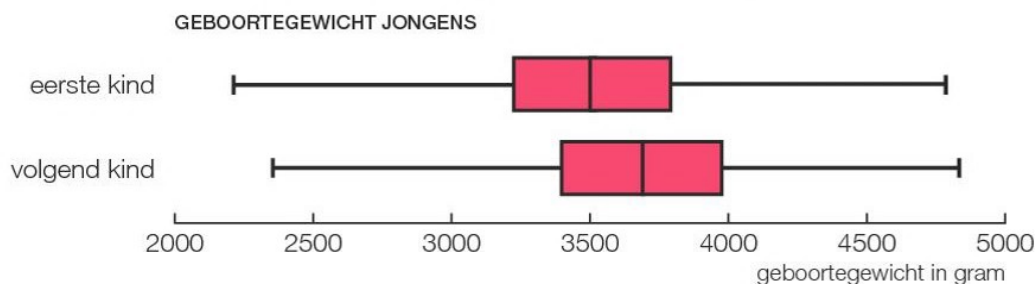
In de tabel zie je de resultaten van een steekproef van jongens die na een normale zwangerschapsduur zijn geboren. Het geboortegewicht van eerste kinderen blijkt gemiddeld lager te zijn dan dat van volgende kinderen.

GEBOORTEGEWICHT JONGENS

	gem.	stand. afw.
eerste kind	3497 gram	428 gram
volgend kind	3673 gram	442 gram

- Onderzoek met een associatiemaat of het verschil in geboortegewicht tussen eerste kind en volgend kind groot, middelmatig of gering is.
- Stel dat de standaardafwijkingen kleiner zouden zijn dan die in de tabel.
Welke invloed heeft dat op de effectgrootte? Licht toe.

Hieronder zijn de boxplots getekend bij de geboortegewichten van jongens die na een normale zwangerschapsduur zijn geboren.



figuur 2.35

- Welke conclusie over het verschil in geboortegewicht kun je trekken op grond van de boxplots?
- Bij vraag a heb je een andere conclusie getrokken dan bij vraag c over de grootte van het verschil in geboortegewicht.
Welke van deze twee conclusies is het beste te verdedigen?
Licht toe.



Terugblik

Associatiematen

Met associatiematen kun je de grootte van het verschil tussen twee groepen kwantificeren, oftewel uitdrukken in een getal. Welke associatiemaat je gebruikt hangt af van het type variabele en de gegevens. Op het centraal examen worden de associatiematen phi, max.Vcp, effectgrootte en het vergelijken van twee boxplots gebruikt. De associatiematen staan samen met de vuistregels voor het bepalen van de grootte van het verschil op het formuleblad. Je hoeft de associatiematen en vuistregels dus niet uit het hoofd te leren. Het is wel belangrijk te weten in welke situatie je welke associatiemaat moet gebruiken.

Verschillen kwantificeren bij nominale variabelen

Voor het kwantificeren van het verschil bij nominale variabelen met twee waarden kun je gebruikmaken van de phi-coëfficiënt.

	waarde I	waarde II	
waarde 1	a	b	$a + b$
waarde 2	c	d	$c + d$
	$a + c$	$b + d$	

$$\text{phi} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

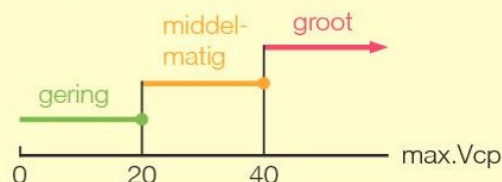
VUISTREGELS PHI-COËFFICIËNT



Verschillen kwantificeren bij ordinale variabelen

Voor het kwantificeren van het verschil bij een ordinale variabele kun je gebruikmaken van het maximale verschil in cumulatief percentage (max.Vcp).

VUISTREGELS MAX.VCP

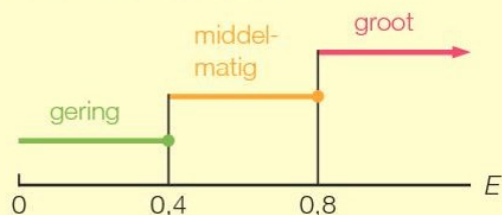


Verschillen kwantificeren bij kwantitatieve variabelen

Bij kwantitatieve variabelen zijn verschillen te kwantificeren met de effectgrootte en door de boxplots te vergelijken.

De formule van de effectgrootte is $E = \frac{\bar{X}_1 - \bar{X}_2}{\frac{1}{2}(S_1 + S_2)}$.

VUISTREGELS EFFECTGROOTTE

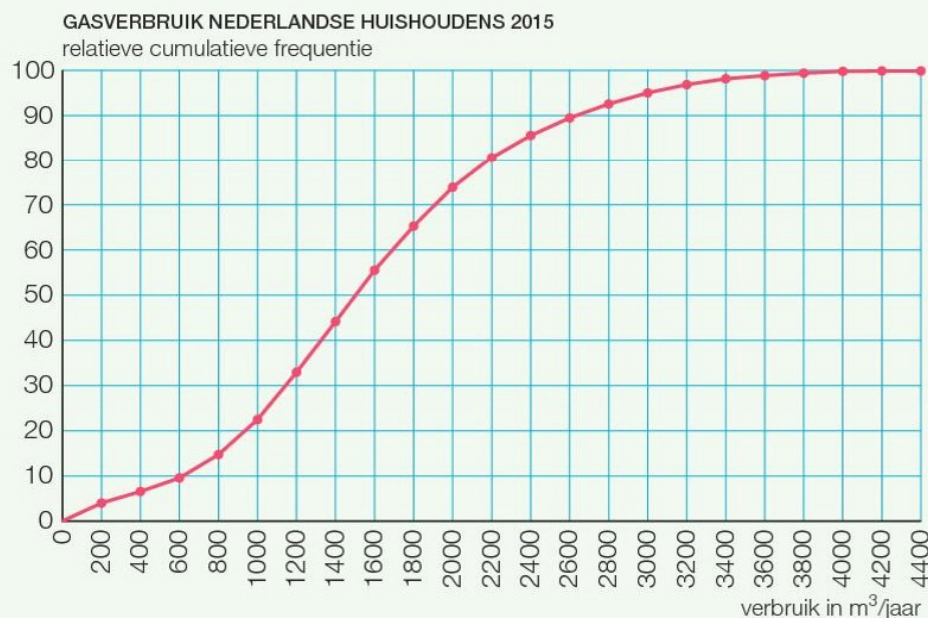


Bij het vergelijken van boxplots gebruik je de volgende vuistregels.

- Als de boxen elkaar niet overlappen, dan zeggen we *het verschil is groot*.
- Als de boxen elkaar wel overlappen en een mediaan van een boxplot ligt buiten de box van de andere boxplot, dan zeggen we *het verschil is middelmatig*.
- In alle andere gevallen zeggen we *het verschil is gering*.

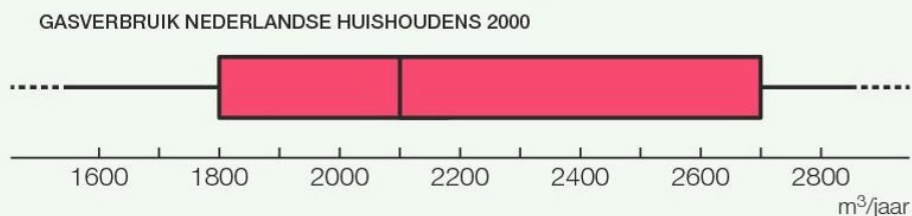
Eindopdracht Nederlands energieverbruik

[▶ WERKBLAD] Hieronder staan twee figuren en een tabel die betrekking hebben op het energieverbruik van Nederlandse huishoudens.



KERNCIJFERS HUISHOUDENS (BRON: RVO)

	2000	2015
aantal inwoners per 1 januari	15.863.950	16.900.726
aantal huishoudens per 1 januari	6.801.000	7.665.000
aantal personen per huishouden	2,3	2,2
gasverbruik per huishouden	1900 m ³	1432 m ³
verbruik elektriciteit per huishouden	3103 kWh	2966 kWh



- Bereken met hoeveel procent het totale elektriciteitsverbruik van Nederlandse huishoudens is veranderd in 2015 ten opzichte van 2000. Doe dit ook voor het gasverbruik.
- Onderzoek of de mediaan van het gasverbruik van Nederlandse huishoudens in 2015 hoger of lager was dan het gemiddelde. Kun je deze conclusie met 100% zekerheid trekken op basis van de gegevens hierboven? Licht toe.
- Onderzoek met behulp van een associatiemaat hoe groot het verschil is in gasverbruik van Nederlandse huishoudens in 2015 vergeleken met dat in 2000.

Diagnostische toets

2.1 De onderzoeksvraag

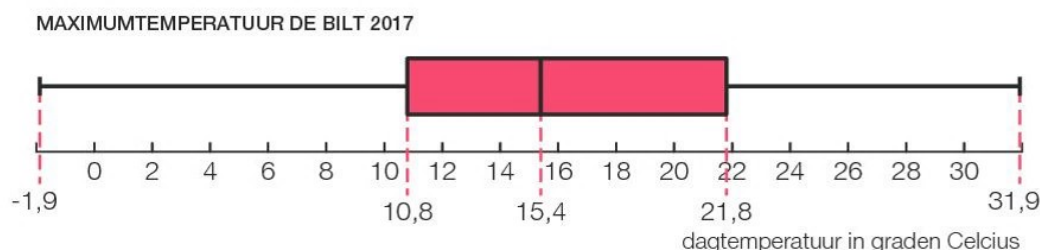
- 1**
- a Welke vier fasen onderscheiden we in de statistische cyclus?
 - b Wat wordt bedoeld met een causaal verband? Geef een voorbeeld.
 - c Aan welke twee eisen moet een representatieve steekproef voldoen?
 - d Wat is het verschil tussen de populatieproportie en de steekproefproportie? Wanneer zullen deze proporties in de meeste gevallen weinig verschillen?
- 2** Van 123 leerlingen van een school met 1447 leerlingen is bekend dat ze meer dan 20 uur per week gamen.
Bij een steekproef geven 12 leerlingen aan dat ze meer dan 20 uur per week gamen, waarmee de steekproefproportie op 0,09 komt.
- a Bereken de populatieproportie van de leerlingen met het kenmerk 'gamt meer dan 20 uur per week'. Rond af op drie decimalen.
 - b Bereken ook het aantal leerlingen in de steekproef.

2.2 Data verzamelen

- 3**
- a Omschrijf het verschil tussen een kwantitatieve en een kwalitatieve variabele. Geef van elk een voorbeeld.
 - b Welke twee meetniveaus ken je? Geef bij elk van de meetniveaus een voorbeeld van een variabele die op dat niveau is gemeten.
 - c Omschrijf het verschil tussen een discrete en een continue variabele. Geef van elk een voorbeeld.

2.3 Centrum- en spreidingsmaten

- 4** In het jaar 2017 is elke dag de maximumtemperatuur in De Bilt genoteerd. Zie de boxplot in figuur 2.35.



figuur 2.35

- 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 Volgens Lars is de spreidingsbreedte meer dan drie keer zo groot als de interkwartielafstand. Onderzoek of Lars gelijk heeft.
- d Geef een schatting van de standaardafwijking. Kies uit 1°C , 4°C , 9°C en 16°C .

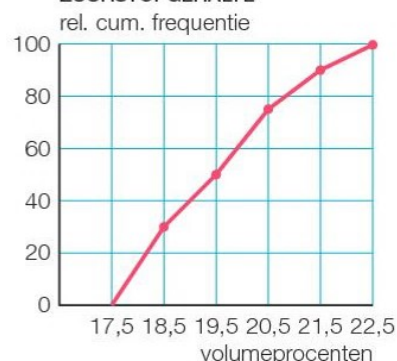
2.4 Representatie van data

- 5** 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.
- 6** 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 frequentietabel met absolute frequenties.
 - c** Teken een histogram met absolute frequenties.

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

ZUURSTOFGEHALTE

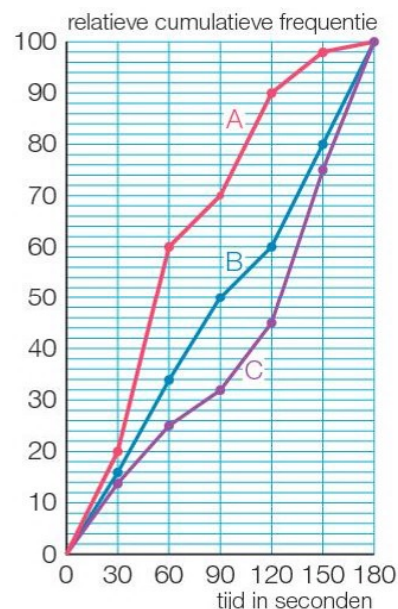


figuur 2.36

	geslaagd	
	ja	nee
A	34	57
B	66	60
	100	117
		217

2.5 Conclusies trekken

- 7** In de tabel hiernaast zie je van het afgelopen jaar hoeveel van de kandidaten van de rijsscholen A en B, die voor het eerst rijexamen deden, slaagden. Onderzoek met de phi-coëfficiënt of het verschil groot, middelmatig of gering is.
- 8** Bij een onderzoek naar de levensduur van espresso-apparaten bleek dat de gemiddelde levensduur van onderzochte apparaten van het type Gustoso 5,4 jaar was en de standaardafwijking 1,6 jaar. Van het type Robusto was het gemiddelde 6,6 jaar en de standaardafwijking 1,2 jaar. Onderzoek of het verschil groot, middelmatig of gering is.
- 9** [▶ WERKBLAD] Een supermarktketen heeft onderzoek gedaan naar de wachttijd bij de kassa. Bij elk filiaal is van 300 klanten de wachttijd genoteerd. In de figuur hiernaast zijn de resultaten van drie filialen verwerkt.
- a** Bij welk filiaal is de wachttijd bij de kassa het kortst?
 - b** Onderzoek met de max.Vcp of het verschil in wachttijd tussen de filialen A en B groot, middelmatig of gering is. Doe dit ook voor de filialen B en C en voor de filialen A en C.
 - c** Teken in één figuur de boxplots van de filialen A, B en C en onderzoek of je met de boxplots dezelfde conclusies trekt als bij vraag b.



figuur 2.37