HOOFDSTUK 2

Onderzoeken en data presenteren

Wiskunde A · havo 5



IN DIT HOOFDSTUK LEER JE:

onderzoeksvragen formuleren en steekproeven beoordelen;
gegevens ordenen en met passende diagrammen presenteren;
centrummaten en spreidingsmaten berekenen en interpreteren.

Wat weet je nog?

1 Een deel van de groep

Van 180 deelnemers kiezen 54 voor een ochtendactiviteit. Bij een nieuwe groep van 240 deelnemers wordt hetzelfde percentage verwacht.

a Hoeveel procent van de 180 deelnemers kiest voor de ochtendactiviteit?

b Hoeveel deelnemers in de nieuwe groep worden dan verwacht bij de ochtendactiviteit?

2 Twee soorten verandering

Het percentage leerlingen dat met de fiets komt, stijgt van 40% naar 52%.

a Hoe groot is de stijging in procentpunten?

b Hoe groot is de relatieve stijging? Geef een percentage.

3 De juiste groep als basis

De tabel geeft het vervoer naar school in twee klassen.

Klas	Lopend	Fiets	Bus
A	6	18	6
B	12	12	6

a Welk percentage van alle leerlingen uit beide klassen fietst?

b Welk percentage van klas B komt lopend?

4 Tijdseenheden

Een tijdmeting geeft 2,4 minuten. Een tweede meting geeft 155 seconden.

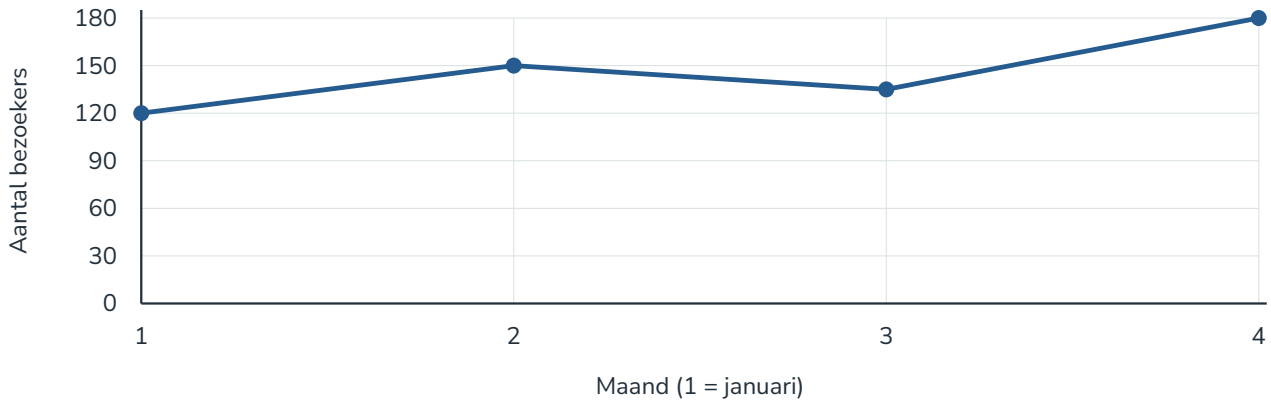
a Schrijf 2,4 minuten in seconden.

b Schrijf 155 seconden in minuten en seconden.

5 Een grafiek lezen

De grafiek toont het aantal bezoekers van een activiteit in vier maanden.

Bezoekers per maand



a In welke maand waren er de meeste bezoekers? Hoeveel waren dat?

b Met hoeveel procent daalt het aantal van februari naar maart?

6 Categorie of hoeveelheid

Op een formulier staan de vier gegevens uit de tabel.

Gegeven	Waarde
Woonwijk	Noord
Aantal busritten deze week	3
Reistijd	18,5 minuten
Oordeel over de busrit	tevreden

a Welke gegevens geven een hoeveelheid aan en welke een categorie?

b Een kluisje heeft nummer 27. Geeft dit nummer een hoeveelheid aan? Leg uit.

2.1 Onderzoeksvragen, steekproeven en soorten gegevens

Leerdoel

Ik kan bij een onderzoeksvraag de populatie afbakenen en de benodigde meetvariabelen kiezen.

Theorie

Een meetbare vraag

Een onderzoeksvraag zegt wat je wilt weten en over wie of wat. De populatie is de volledige groep waarover je een uitspraak wilt doen. Baken die groep af met bijvoorbeeld plaats, leeftijd en periode. Kies ook de onderzoekseenheid: één persoon, één rit of één bezoek. Dezelfde persoon kan meerdere bezoeken afleggen; dat zijn dan meerdere onderzoekseenheden.

Een variabele is een kenmerk dat je bij iedere onderzoekseenheid vastlegt. Kies gegevens waarmee je de vraag werkelijk kunt beantwoorden. Leg bij een meting de eenheid en het meetmoment vast. Maak woorden als «lang» of «vroeg» meetbaar met een grens. Let op het verschil tussen «meer dan», «minstens» en «vóór». Soms kun je een benodigde waarde uit twee geregistreerde gegevens terugrekenen.

Zie voorbeeld 2.1A, blz. 211.

Onthoud: Baken de groep af en kies gegevens die de onderzoeksvraag beantwoorden.

1 Langer dan een halfuur

Een school onderzoekt: «Welk percentage van de huidige leerlingen reist op een gewone lesdag langer dan 30 minuten van huis naar school?» Er zijn 960 leerlingen. In een kleine proefmeting zijn de reistijden 18, 24, 30, 36 en 42 minuten.

a Welke personen horen bij de populatie? Gebruik de afbakening in de onderzoeksvraag.

b Welke variabele moet per leerling worden gemeten? Noem ook de eenheid.

c Welk percentage van de vijf gemeten leerlingen reist langer dan 30 minuten?

2 Wachten aan de balie

Een bibliotheek wil weten welk deel van de baliebezoeken op één afgesproken zaterdag lang moet wachten. Het onderzoek gaat over bezoeken waarbij iemand tussen 09:00 uur en 12:00 uur in de rij aansluit; 12:00 uur zelf valt erbuiten. «Lang» betekent meer dan 5 minuten. De tabel geeft drie proefmetingen.

Bezoek	Aansluiten	Begin bediening
1	09:10	09:12
2	09:25	09:31
3	11:54	12:03

a Formuleer een precieze onderzoeksvraag en benoem de populatie.

b Bereken de drie wachttijden. Welk percentage van deze bezoeken voldoet aan «lang wachten»? Rond af op één decimaal.

3 Vroeg vertrekken

Alle 25 leerlingen van een havo 5-klas registreren op woensdag hun aankomsttijd op school en hun reistijd van huis naar school. De school vraagt: «Welk percentage van deze klas vertrekt die woensdag vóór 07:00 uur van huis?» De leerlingen hebben geen tussenstop.

Aantal leerlingen	Aankomsttijd	Reistijd
8	08:15	75 minuten
7	08:05	70 minuten
10	08:10	50 minuten

a Welke gegevens uit de tabel heb je nodig om de vraag te beantwoorden? Leg uit hoe je daarmee de vertrektijd vindt.

b Bereken het gevraagde percentage en leg uit hoe je de grens van 07:00 uur gebruikt.

Wie levert de gegevens?

Leerdoel

Ik kan beoordelen hoe een steekproef is geselecteerd en uitleggen waardoor een uitkomst vertekend kan zijn.

Theorie

Wie levert de gegevens?

Vaak onderzoek je een deel van de populatie: een steekproef. De steekproefomvang is het aantal onderzochte eenheden. Bij een aselecte selectie heeft iedere eenheid uit de populatie dezelfde kans om te worden gekozen. Loting uit een volledige lijst kan zo'n selectie opleveren. Zelf aanmelden of alleen aanwezige bezoekers vragen geeft die gelijke kans meestal niet.

Een representatieve steekproef weerspiegelt de populatie op kenmerken die voor de onderzoeksvraag van belang zijn. Aselect kiezen helpt, maar garandeert dit niet. Ook een grote steekproef kan vertekend zijn. Non-respons ontstaat als geselecteerde personen niet antwoorden. Hun mening of gedrag kan anders zijn dan die van respondenten. Onderscheid daarom uitgenodigden en respondenten. Bereken een percentage met de groep die daadwerkelijk is gemeten als basis.

Zie voorbeeld 2.1A, blz. 211.

Onthoud: Aselect kiezen en veel antwoorden garanderen geen representatieve respondentengroep.

4 Een link bij de training

Een sportclub heeft 600 leden. Tijdens de avondtraining hangt een QR-code voor een enquête over de trainingstijden. Er reageren 60 leden; 36 van hen willen een extra ochtendtraining.

a Wat zijn hier de populatie, de steekproef en de steekproefomvang?

b Is de selectie aselect? Noem twee manieren waarop de uitkomst kan worden vertekend.

c Bereken het percentage voor de respondenten. Hoe kan de club de selectie verbeteren?

5 Wel geloot, niet allemaal antwoord

Een theater heeft 500 abonnementhouders. Het loot 80 verschillende personen uit de volledige lijst; ieder heeft dezelfde kans. Van hen antwoorden 55 personen op een enquête. Van die 55 zijn er 44 tevreden over de nieuwe begintijd.

a Beoordeel de selectie van de 80 uitgenodigden en die van de 55 respondenten. Leg het verschil uit.

b Bereken het percentage tevreden respondenten. Een medewerker schrijft: «80% van alle abonnementhouders is tevreden.» Beoordeel dit en geef één gerichte verbetering.

Wat voor gegeven is het?

Leerdoel

Ik kan een variabele indelen als nominaal, ordinaal, discreet kwantitatief of continu kwantitatief en mijn keuze uitleggen.

Theorie

Wat voor gegeven is het?

Een kwalitatieve variabele beschrijft een categorie. Bij een nominale variabele is er geen natuurlijke rangorde, zoals bij vervoermiddel. Bij een ordinale variabele is die rangorde er wel, zoals bij ontevreden, neutraal en tevreden. Cijfers kunnen categoriecodes zijn. De code 3 is dan geen hoeveelheid; een gemiddelde van zulke codes heeft zonder verdere motivatie geen betekenis.

Een kwantitatieve variabele geeft een hoeveelheid aan. Een discrete variabele heeft afzonderlijke mogelijke waarden; een telling zoals het aantal bezoeken geeft hele getallen. Een continue variabele is een meting die ook tussenwaarden kan aannemen, zoals afstand of tijdsduur. Kijk naar de onderliggende variabele. Een temperatuur die op hele graden wordt afgerond, blijft continu. Vraag bij een getal dus eerst of het iets aanduidt, telt of meet.

Zie voorbeeld 2.1B, blz. 213.

Onthoud: Bepaal het soort van de variabele, niet het soort van de afgedrukte waarde.

6 Vier soorten variabelen

Een vragenlijst over wonen registreert de kenmerken in de tabel.

Variabele	Mogelijke registratie
Woningtype	appartement
Tevredenheid	ontevreden / neutraal / tevreden
Aantal huisgenoten	2
Reisafstand naar werk	8,7 km

a Welke twee variabelen zijn kwalitatief? Geef voor beide aan of ze nominaal of ordinaal zijn.

6 Vier soorten variabelen (vervolg)

Een vragenlijst over wonen registreert de kenmerken in de tabel.

Variabele	Mogelijke registratie
Woningtype	appartement
Tevredenheid	ontevreden / neutraal / tevreden
Aantal huisgenoten	2
Reisafstand naar werk	8,7 km

b De andere twee variabelen zijn kwantitatief. Welke is discreet en welke continu? Licht beide keuzes toe.

c Het formulier registreert ook woningnummer 204. Waarom is dit geen kwantitatieve variabele?

7 Hetzelfde getal, een ander soort

Een reparatiewerkplaats registreert de variabelen in de tabel.

Variabele	Registratie
Type apparaat	telefoon / tablet / laptop
Urgentie	laag / middel / hoog
Aantal vervangen onderdelen	een geheel aantal
Reparatieduur	afgerond op hele minuten

a Deel de vier variabelen in als nominaal, ordinaal, discreet kwantitatief of continu kwantitatief. Leg iedere keuze kort uit.

b Op één regel staan klantnummer 104 en aantal bestellingen 104. Leg uit waarom deze gegevens verschillende soorten variabelen zijn.

8 Een code met een betekenis

Vijf bezoekers antwoorden op de vraag «Gebruik je een leenfiets?». Nee krijgt code 0 en ja code 1. De geregistreerde codes zijn 1, 0, 1, 1 en 0.

a Welk soort variabele is het antwoord op deze vraag? Verandert de codering dat?

b Het gemiddelde van de codes is 0,6. Leg uit waarom dit getal hier toch een bruikbare betekenis heeft.

Van gegevens naar conclusie

Leerdoel

Ik kan de stappen van een onderzoek herkennen en een conclusie formuleren die past bij de verzamelde gegevens.

Theorie

Van gegevens naar conclusie

Een onderzoek doorloopt vier stappen. Bij onderzoeksvraag bepaal je wat je wilt weten, over welke populatie en in welke periode. Bij data verzamelen kies je een passende selectie en registreer je de benodigde variabelen. Bij data analyseren orden je de gegevens, vergelijk je groepen en bereken je bijvoorbeeld percentages. Bij conclusies trekken beantwoord je de vraag met de uitkomsten en hun beperkingen.

Noem in een conclusie welke groep en periode de gegevens beschrijven. Een uitkomst bij respondenten is niet automatisch de uitkomst van de hele populatie. Een verband tussen variabelen betekent ook niet dat het ene het andere veroorzaakt. Groepen kunnen op andere kenmerken verschillen. Controleer of de berekening, de selectie en de meetwijze je uitspraak ondersteunen. Een onverwachte uitkomst kan aanleiding zijn voor een nieuwe onderzoeksvraag.

Zie voorbeeld 2.1B, blz. 213.

Onthoud: Een conclusie mag alleen zo ver gaan als de gegevens en de onderzoeksopzet toelaten.

9 Een te sterke conclusie

Van 120 havo 5-leerlingen van één school reageren 40 vrijwilligers op een enquête over ontbijt en op tijd komen. De tabel beschrijft deze 40 respondenten. Een leerling concludeert: «Ontbijten zorgt dat havo 5-leerlingen vaker op tijd komen.»

Ontbijt die ochtend	Op tijd	Te laat
Wel	21	7
Niet	9	3

a Bereken binnen beide ontbijtgroepen het percentage dat op tijd komt. Ondersteunt de tabel dat «vaker»?

9 Een te sterke conclusie (vervolg)

Van 120 havo 5-leerlingen van één school reageren 40 vrijwilligers op een enquête over ontbijt en op tijd komen. De tabel beschrijft deze 40 respondenten. Een leerling concludeert: «Ontbijten zorgt dat havo 5-leerlingen vaker op tijd komen.»

Ontbijt die ochtend	Op tijd	Te laat
Wel	21	7
Niet	9	3

b Noem naast «vaker» twee andere problemen met de conclusie. Formuleer een conclusie die wel door de tabel wordt gedragen.

10 De volgorde van het onderzoek

Een school onderzoekt een aanpassing van de pauzetijd. Vier handelingen staan in willekeurige volgorde.

Handeling	Beschrijving
P	De 24 leerlingen van één havo 5-klas beantwoorden een vragenlijst.
Q	In deze klas heeft twee derde voorkeur voor een langere pauze.
R	Welk percentage leerlingen van deze klas wil een langere pauze?
S	Er worden 16 voorstanders en 8 tegenstanders geteld.

a Zet de handelingen in de volgorde van de onderzoekscyclus. Noem bij iedere handeling de stap.

b De directie wil met deze gegevens besluiten over alle leerlingen. Welke beperking moet zij meenemen, en welke vervolgstap past daarbij?

11 Apps en een abonnement

Een leesclub heeft 200 leden. Uit de volledige ledenlijst worden 48 leden geloot, ieder met dezelfde kans. Alle 48 doen mee. De tabel geeft aan of zij een nieuwsapp gebruiken en een digitaal nieuwsabonnement hebben.

Nieuwsapp	Wel abonnement	Geen abonnement
Wel	12	6
Niet	9	21

a Welk soort variabele zijn nieuwsapp gebruiken en nieuwsabonnement hebben? Leg uit.

b Bereken in beide appgroepen het percentage met een abonnement. Hoe groot is het verschil in procentpunten? Rond af op één decimaal.

c Formuleer een conclusie over de 48 deelnemers. Beoordeel daarnaast de uitspraken «Een app veroorzaakt meer abonnementen» en «De steekproef is zeker representatief voor de 200 leden».

12 Een eigen onderzoeksopzet

Een gemeentelijk museum wil weten welk percentage van alle 14- tot en met 18-jarige inwoners in september minstens één museum heeft bezocht. Via het Instagramaccount van het museum reageren 80 jongeren; 52 melden minstens één bezoek.

a Welk gegeven zou je per jongere registreren om de vraag te beantwoorden? Kies zelf een geschikte variabele en benoem het soort.

b Ontwerp een selectie die beter bij de populatie past. Benoem ook een probleem dat na de selectie nog kan ontstaan.

c Bereken het percentage in de huidige enquête en schrijf een conclusie die precies aangeeft welke groep is beschreven.

13 Fietsen naar het werk

Een werkgever heeft 240 medewerkers, inclusief thuiswerkers. De onderzoeksvraag is: «Welk percentage medewerkers fietste in de afgelopen werkweek op minstens twee dagen naar het werk?» In de fietsenstalling worden 60 medewerkers benaderd. Er antwoorden 46; 23 van hen fietsten op minstens twee dagen naar het werk.

a Benoem de populatie en een variabele waarmee de onderzoeksvraag kan worden beantwoord.

b Bereken de uitkomst voor de respondenten. Kan dit worden gepresenteerd als het percentage voor alle 240 medewerkers? Licht je oordeel toe.

14 Digitaal lenen en leestijd

Een bibliotheek loot 60 leden uit de volledige lijst van 300 leden; ieder heeft dezelfde kans. Alle 60 antwoorden. De eerste tabel geeft de geregistreerde variabelen. De tweede tabel vat digitaal lenen en een leestijd van meer dan 30 minuten per dag samen.

Variabele	Registratie
Favoriet genre	bijvoorbeeld roman of reisboek
Tevredenheid over het aanbod	ontevreden / neutraal / tevreden
Aantal digitale leningen afgelopen maand	een geheel aantal
Leestijd per dag	afgerond op hele minuten

Digitale lener	Meer dan 30 minuten	Hoogstens 30 minuten
Wel	18	6
Niet	18	18

a Deel de vier variabelen uit de eerste tabel in als nominaal, ordinaal, discreet kwantitatief of continu kwantitatief. Licht je keuzes kort toe.

b Bereken met de tweede tabel binnen beide groepen het percentage dat meer dan 30 minuten per dag leest. Bij welke stap van de onderzoekscyclus hoort deze berekening?

c Formuleer een conclusie die de tweede tabel ondersteunt. Leg uit of de uitspraak «Digitaal lenen zorgt dat leden langer lezen» gerechtvaardigd is.

2.2 Frequentietabellen en klassenindeling

Leerdoel

Ik kan ruwe gegevens ordenen in een frequentietabel met correcte totalen.

Theorie

Van gegevens naar aantallen

Ruwe gegevens zijn de afzonderlijke waarnemingen, bijvoorbeeld het aantal bezoeken per persoon. In een frequentietabel krijgt iedere verschillende waarde of categorie één rij. De absolute frequentie is het aantal keer dat die waarde of categorie voorkomt. Zet getalswaarden van klein naar groot; kies bij categorieën een duidelijke volgorde, bijvoorbeeld alfabetisch. Als een mogelijke waarde niet voorkomt, kun je daarvoor een rij met frequentie 0 opnemen.

Werk de waarnemingen één voor één af en gebruik zo nodig turfstreepjes. Tel iedere waarneming precies eenmaal, ook een gemeten waarde 0. Tel daarna de frequenties op. Deze som moet gelijk zijn aan het totale aantal waarnemingen. Een totaalcontrole kan een vergeten of dubbel getelde waarneming zichtbaar maken, maar bewijst nog niet dat iedere rij klopt. Controleer bij twijfel ook de telling per waarde.

Zie voorbeeld 2.2A, blz. 215.

Onthoud: De som van de absolute frequenties is het totale aantal waarnemingen.

1 Een totaal controleren

Een inventarisatie bevat de volgende aantallen boeken.

Vorm	Aantal
Papier	7
Digitaal	5
Audio	4

a Hoeveel boeken zijn in de tabel opgenomen?

b Er moesten 20 boeken worden geregistreerd. Hoeveel boeken ontbreken nog in de tabel?

2 Een ontbrekend aantal

Alle 28 deelnemers kiezen precies één werkplek. Eén aantal is weggevallen.

Werkplek	Aantal deelnemers
A	9
B	8
C	?
D	6

a Bereken het ontbrekende aantal voor werkplek C.

b Heeft precies de helft van de deelnemers werkplek B of D gekozen?

3 Tellen zonder overslaan

Oefengegevens: 18 deelnemers registreren hoeveel online oefensessies zij op één dag starten. De waarden zijn: 0, 1, 2, 1, 0, 3, 2, 1, 0, 2, 3, 1, 2, 0, 1, 3, 2, 1.

a Maak een frequentietabel voor 0, 1, 2 en 3 gestarte sessies. Werk de lijst van links naar rechts af.

b Controleer het totaal. Waarom mag je de waarden 0 bij het tellen niet overslaan?

4 Een eigen frequentietabel

Oefengegevens: het aantal bezoeken aan een uitleenpunt is voor 20 gebruikers genoteerd: 0, 2, 1, 3, 2, 4, 1, 2, 0, 1, 3, 2, 1, 4, 2, 0, 3, 1, 2, 2.

a Orden de gegevens in een frequentietabel en controleer het totaal.

b Hoeveel gebruikers bezochten het uitleenpunt minstens drie keer?

5 Categorieën ordenen

Oefengegevens: 16 deelnemers kiezen één fruitsmaak. A staat voor appel, C voor citroen en P voor peer. De keuzes zijn: A, C, C, P, A, P, C, A, C, P, P, A, C, C, A, P.

a Maak een frequentietabel met de volledige namen van de smaken.

b Je wilt de populairste smaak bovenaan zetten. Welke rij komt dan eerst? Verandert daardoor het totaal?

Aantallen en percentages

Leerdoel

Ik kan absolute, relatieve en cumulatieve frequenties berekenen en interpreteren.

Theorie

Aantallen en percentages

De relatieve frequentie geeft aan welk deel van alle waarnemingen bij een waarde of categorie hoort. Deel de absolute frequentie door het totale aantal waarnemingen. Voor een percentage vermenigvuldig je de uitkomst met 100. Vermeld altijd op welke groep het percentage betrekking heeft. Verandert het totaal, dan kan het percentage veranderen, zelfs als het betreffende aantal gelijk blijft. Reken met ongeronde aantallen en rond pas het uiteindelijke percentage af.

Bij natuurlijk geordende waarden kun je frequenties cumuleren: tel de frequentie van een rij op bij die van alle eerdere rijen. De cumulatieve absolute frequentie bij waarde x telt alle waarnemingen met waarde hoogstens x. De laatste cumulatieve frequentie is het totale aantal; het laatste cumulatieve percentage is 100%. Bij ongeordende categorieën heeft deze optelling geen inhoudelijke betekenis. Door afronding kunnen losse percentages samen iets van 100% afwijken.

Zie voorbeeld 2.2A, blz. 215.

Onthoud: Relatief: delen door het totaal; cumulatief: optellen tot en met de gekozen waarde.

6 Een tabel uitbreiden

Oefengegevens: voor 20 deelnemers is het aantal bezochte workshops geteld.

Aantal workshops	Absolute frequentie
0	3
1	6
2	7
3	3
4	1

a Voeg een kolom relatieve frequentie in procenten toe. Deel elk aantal door 20.

6 Een tabel uitbreiden (vervolg)

Oefengegevens: voor 20 deelnemers is het aantal bezochte workshops geteld.

Aantal workshops	Absolute frequentie
0	3
1	6
2	7
3	3
4	1

b Voeg de cumulatieve absolute frequenties toe. Welk percentage bezocht hoogstens twee workshops?

7 Twee verschillende grenzen

Oefengegevens: voor 20 straten is het aantal pakketbezorgingen op één ochtend genoteerd.

Aantal bezorgingen	Absolute frequentie
1	4
2	8
3	6
4	2

a Bereken het percentage straten met precies twee bezorgingen.

b Bereken het percentage straten met meer dan twee bezorgingen.

c Geef de cumulatieve absolute frequentie bij drie bezorgingen en leg uit wat dit aantal betekent.

8 Terug van procenten naar aantallen

Oefengegevens: 40 deelnemers geven aan hoeveel keren zij deze week een deelauto gebruikten.

Aantal keren	Relatieve frequentie
0	15%
1	30%
2	40%
3	15%

a Bereken de absolute frequenties.

b Hoeveel deelnemers gebruikten de deelauto hoogstens één keer?

9 Een percentage verantwoorden

Oefengegevens: op 24 dagen is het aantal teruggebrachte kratten geregistreerd: 0, 2, 1, 3, 2, 0, 1, 2, 4, 3, 1, 2, 0, 2, 1, 3, 2, 4, 1, 0, 2, 3, 1, 2.

a Maak één frequentietabel waarmee je zowel percentages als aantallen tot en met elke waarde kunt aflezen. Rond percentages af op één decimaal.

b Op welk percentage van deze dagen werden minstens twee kratten teruggebracht? Rond af op één decimaal.

c Er worden twee extra dagen toegevoegd, beide met één teruggebracht krat. Bereken opnieuw het percentage dagen met minstens twee kratten. Rond af op één decimaal.

10 Cumuleren vraagt een volgorde

Oefengegevens: 40 bezoekers beoordelen een wachtruimte met de geordende categorieën slecht, matig, goed en uitstekend.

Beoordeling	Absolute frequentie
Slecht	8
Matig	12
Goed	15
Uitstekend	5

a Voeg de relatieve en cumulatieve relatieve frequenties in procenten toe.

b Een ander zet de rij met de hoogste frequentie bovenaan en cumuleert dan in de volgorde goed, matig, slecht, uitstekend. Waarom kun je de som bij matig dan niet lezen als het percentage met beoordeling hoogstens matig?

Gegevens indelen in klassen

Leerdoel

Ik kan een klassenindeling maken met consistente grenzen en een passende klassebreedte.

Theorie

Gegevens indelen in klassen

Bij veel verschillende getalswaarden kun je meerdere waarden samenvoegen in klassen. Kies een startgrens en een klassebreedte: het verschil tussen de boven- en ondergrens. Met startgrens 10 en breedte 5 ontstaan bijvoorbeeld $10 \leq x < 15$, $15 \leq x < 20$ en $20 \leq x < 25$. De ondergrens telt steeds mee, de bovengrens niet. Een waarde 15 hoort dus bij de tweede klasse. Maak genoeg klassen en kies de laatste bovengrens boven de grootste waarneming.

De klassen moeten op elkaar aansluiten zonder gaten of overlap, zodat elke waarneming in precies één klasse valt. Tel daarna per klasse en controleer de totale frequentie. Smalle klassen tonen meer detail, maar kunnen veel lege of kleine groepen opleveren. Brede klassen geven meer overzicht, maar verbergen verschillen. Kies de breedte passend bij het doel. Een afgerond geregistreerde meting blijft een continue variabele.

Zie voorbeeld 2.2B, blz. 217.

Onthoud: Gebruik aansluitende grenzen: de ondergrens telt mee en de bovengrens niet.

11 Een waarde op een grens

Oefengegevens: wachttijden t in seconden zijn 18,6; 21,4; 24,8; 25,0; 26,1; 29,9; 30,0; 32,7; 34,2; 35,0; 37,8; 39,6. Gebruik de klassen $15 \leq t < 20$, $20 \leq t < 25$, $25 \leq t < 30$, $30 \leq t < 35$ en $35 \leq t < 40$.

a In welke klasse hoort 25,0 seconden? En in welke klasse hoort 30,0 seconden?

b Maak de frequentietabel en controleer het totaal.

12 Zelf klassegrenzen maken

Oefengegevens: afstanden d in kilometer zijn 3,8; 4,0; 4,6; 5,1; 5,9; 6,0; 6,3; 7,2; 7,9; 8,0; 8,4; 9,5; 9,8; 10,0; 11,6; 11,9.

a Maak een frequentietabel met startgrens 2 km en klassebreedte 2 km. Kies de laatste bovengrens boven de grootste meting.

b Leg met de klassegrenzen uit waarom 8,0 km niet bij de klasse $6 \leq d < 8$ hoort.

c Welk percentage van de afstanden is kleiner dan 8 km? Rond af op één decimaal.

13 Een breedte kiezen

Oefengegevens: reactietijden t van een digitaal systeem, in milliseconden, zijn 28; 31; 34; 36; 39; 41; 44; 47; 49; 50; 53; 55; 58; 61; 63; 65; 68; 70; 74; 78; 80; 83; 87; 89.

a Kies een klassebreedte van 5, 10 of 20 ms. Gebruik startgrens 20 ms, gelijke breedtes en minstens vijf maar hoogstens acht klassen. Alle waarden moeten passen. Leg je keuze uit.

b Maak de frequentietabel bij je keuze.

c Een eis is: minstens 55% van deze reactietijden is korter dan 60 ms. Voldoen deze gegevens?

Wat een klassentabel verbergt

Leerdoel

Ik kan uitleggen welke informatie verloren gaat bij het groeperen van gegevens.

Theorie

Wat een klassentabel verbergt

Een klassentabel bewaart het aantal waarnemingen per interval, maar niet de afzonderlijke waarden. Bij een frequentie 6 in $10 \leq t < 20$ weet je dat zes waarden ten minste 10 en kleiner dan 20 zijn. Je weet niet welke zes waarden dat waren. Verschillende ruwe datasets kunnen daarom dezelfde klassentabel opleveren.

Een grens die samenvalt met een klassegrens kun je vaak exact beoordelen: voor $t < 20$ tel je alle klassen vóór 20 op. Ligt een grens midden in een klasse, dan ontbreekt informatie. Voor $t < 15$ tel je de waarden vóór 10 zeker mee; van de zes waarden tussen 10 en 20 kunnen er nul tot zes bijkomen. Geef dan een kleinste en grootste mogelijk aantal. Bij een vraag om een exact aantal heb je de ruwe gegevens of fijnere informatie nodig.

Zie voorbeeld 2.2B, blz. 217.

Onthoud: Bij een grens binnen een klasse is het exacte aantal meestal niet meer af te lezen.

14 Exact of alleen begrensd?

Oefengegevens: 20 afhandeltijden t in minuten zijn gegroepeerd in de volgende tabel.

Afhandeltijd t (min)	Absolute frequentie
$0 \leq t < 10$	4
$10 \leq t < 20$	7
$20 \leq t < 30$	5
$30 \leq t < 40$	4

a Hoeveel afhandeltijden zijn kleiner dan 20 minuten? Leg uit waarom dit aantal exact af te lezen is.

b Hoeveel afhandeltijden kunnen kleiner dan 15 minuten zijn? Geef het kleinste en grootste mogelijke aantal.

14 Exact of alleen begrensd? (vervolg)

Oefengegevens: 20 afhandeltijden t in minuten zijn gegroepeerd in de volgende tabel.

Afhandeltijd t (min)	Absolute frequentie
$0 \le t < 10$	4
$10 \le t < 20$	7
$20 \le t < 30$	5
$30 \le t < 40$	4

c Noem twee mogelijke waarden voor één afhandeltijd uit $20 \le t < 30$. Kan de tabel beslissen welke van die waarden gemeten is?

15 Twee lijsten bij één tabel

Deze oefentabel groepeert 12 duurmetingen t in minuten. De waarden zijn als hele minuten geregistreerd.

Duur t (min)	Absolute frequentie
$0 \le t < 5$	3
$5 \le t < 10$	5
$10 \le t < 15$	4

a Stel twee verschillende lijsten van 12 waarden samen die beide de tabel opleveren. In lijst A moeten precies drie waarden kleiner dan 7 zijn; in lijst B precies acht.

b Is in elke mogelijke lijst precies 25% van de waarden kleiner dan 5? Leg uit.

16 Een rapport controleren

Oefengegevens: 20 verwerkingstijden t in seconden zijn gegroepeerd in de tabel. Een rapport bevat drie uitspraken: I. Precies 15 tijden zijn korter dan 20 s. II. Precies 15 tijden zijn korter dan 18 s. III. Een tijd van 20 s hoort bij $10 \leq t < 20$.

Verwerkingstijd t (s)	Absolute frequentie
$0 \leq t < 10$	6
$10 \leq t < 20$	9
$20 \leq t < 30$	5

a Beoordeel iedere uitspraak. Verbeter onjuiste uitspraken en onderbouw je beoordeling.

b Is de tabel voldoende om te beslissen of minstens 60% van de tijden korter dan 18 s is? Onderbouw met twee mogelijke aantallen.

17 Dagregistraties

Oefengegevens: op 25 dagen is het aantal helpdeskvragen geregistreerd: 2, 1, 0, 3, 2, 1, 4, 2, 0, 1, 3, 2, 1, 2, 0, 3, 1, 4, 2, 1, 3, 2, 0, 1, 2.

a Maak een frequentietabel met absolute, relatieve en cumulatieve absolute frequenties.

b Controleer de uitspraak: op minstens driekwart van deze dagen waren er hoogstens twee helpdeskvragen.

c Twee vergeten dagen blijken allebei twee helpdeskvragen te hebben gehad. Welk percentage van alle 27 dagen had twee vragen? Rond af op één decimaal.

18 Een klassentabel gebruiken

Oefengegevens: lengtes l in centimeter zijn 10,8; 13,2; 14,9; 15,0; 15,7; 16,9; 18,1; 19,9; 20,0; 20,6; 21,5; 22,8; 23,4; 24,9; 25,0; 28,6.

a Maak een frequentietabel met startgrens 10 cm en klassebreedte 5 cm. De laatste bovengrens ligt boven de grootste meting.

b Een klasgenoot krijgt alleen jouw klassentabel. Kan die precies bepalen hoeveel lengtes kleiner dan 22 cm zijn? Geef het kleinste en grootste mogelijke aantal.

c Bereken uit de oorspronkelijke lijst welk percentage van de lengtes kleiner dan 22 cm is. Rond af op één decimaal.

2.3 Diagrammen kiezen en omzetten

Leerdoel

Ik kan een diagram kiezen dat past bij het gegevenstype en de onderzoeksvraag.

Theorie

Kies vanuit de onderzoeksvraag

Begin met de onderzoeksvraag en het gegevenstype. Categorieën vergelijk je met losse staven. Een cirkeldiagram toont aandelen van één geheel; iedere waarneming moet in precies één categorie vallen. Bij ordinale categorieën behoud je de rangorde. Kwantitatieve waarden kun je afzonderlijk tonen in een dotplot of steelbladdiagram. Voor een gegroepeerde continue variabele passen een histogram of frequentiepolygoon.

Een lijndiagram laat een ontwikkeling langs opeenvolgende tijdstippen zien. Twee kwantitatieve gegevens van dezelfde persoon kun je als één punt in een puntenwolk zetten. Twee categorieën per persoon passen in een kruistabel. Wil je een aandeel onder een grens aflezen, kies dan een cumulatieve voorstelling. Cumuleren vereist een natuurlijke ordening. Vraag bij iedere keuze welke informatie zichtbaar moet blijven.

Zie voorbeeld 2.3A, blz. 219.

Onthoud: Kies de voorstelling bij het gegevenstype én de vraag.

1 Aantallen en basis

Twintig leerlingen kiezen één leesvorm. Deze onderzoeksgegevens zijn samengesteld.

Leesvorm	Aantal
Papier	7
Tablet	9
Beide	4

a Bepaal het percentage leerlingen dat een tablet kiest.

b Welk diagram past als je de drie aantallen wilt vergelijken? Licht je keuze toe.

2 Volgorde herkennen

Een beoordelingslijst gebruikt de antwoorden onvoldoende, voldoende en goed. Deze onderzoeksgegevens zijn samengesteld.

Beoordeling	Aantal
Onvoldoende	3
Voldoende	8
Goed	9

a Is deze variabele nominaal of ordinaal? Leg uit.

b Je maakt een staafdiagram. In welke volgorde zet je de categorieën? Mag je ze verbinden om een ontwikkeling in de tijd te tonen?

3 De vraag bepaalt de keuze

Vijfentwintig bezoekers kiezen één activiteit. Deze onderzoeksgegevens zijn samengesteld.

Activiteit	Aantal
Zwemmen	12
Klimmen	8
Dansen	5

a Je wilt zien welke activiteit het vaakst gekozen is. Kies uit een staaf-, cirkel- of lijndiagram. Benoem het gegevenstype en leg je keuze uit.

b Je wilt juist de verdeling van alle 25 keuzes als één geheel tonen. Welk diagram past? Bereken het aandeel zwemmen.

c Waarom voegt het verbinden van de drie aantallen tot een lijn geen zinvolle informatie toe?

4 Zelf een voorstelling kiezen

Kies telkens een voorstelling die de genoemde vraag beantwoordt. De gegevens zijn voor deze oefeningen samengesteld.

a Een bibliotheek telt de bezoekers per maand. Je wilt de ontwikkeling gedurende één jaar zien. Welke voorstelling kies je? Benoem beide assen.

b Van 30 leerlingen zijn per leerling reistijd en afstand tot school bekend. Je wilt nagaan hoe beide samenhangen. Welke voorstelling kies je? Welke brongegevens moet je bij elkaar houden?

c Van dezelfde leerlingen is het vervoermiddel bekend. Je wilt de aandelen fietsen, lopen en bus tonen. Welke voorstelling kies je en wat is de percentagebasis?

5 Welke informatie blijft zichtbaar?

Een onderzoeker bewaart 40 afzonderlijke wachttijden en de gekozen betaalwijze per bezoeker. De gegevens zijn samengesteld.

a Je wilt alle afzonderlijke wachttijden én herhaalde waarden zichtbaar houden. Noem twee passende voorstellingen en leg uit wat beide bewaren.

b Je wilt aflezen welk aandeel korter dan een gekozen grens wacht. Welke voorstelling is dan doelmatig? Waarom is hetzelfde idee niet zinvol voor de betaalwijzen contant, pin en telefoon?

Van frequentietabel naar diagram

Leerdoel

Ik kan een frequentietabel omzetten in een staaf-, cirkel- of lijndiagram en een frequentiepolygoon.

Theorie

Van frequentietabel naar diagram

Een staafdiagram geeft categorieën met losse staven van gelijke breedte. De staafhoogte is de frequentie; begin de verticale as bij 0. Een cirkeldiagram verdeelt 360° over één geheel. Bereken een sectorhoek met frequentie / totaal × 360°. Het percentage is frequentie / totaal × 100%. Controleer dat de sectoren samen 360° zijn.

Een lijndiagram zet op ieder tijdstip de gemeten waarde en verbindt opeenvolgende punten. Bij een continue variabele gebruik je eenduidige klassen, bijvoorbeeld $10 \le t < 20$; vermeld of de laatste bovengrens meetelt. In een histogram met gelijke klassebreedten sluiten de staven aan en is de hoogte de frequentie. Een frequentiepolygoon zet die frequentie bij het klassenmidden: (ondergrens + bovengrens) / 2. Verbind de punten. Geef iedere as een naam, eenheid en passende schaal.

Zie voorbeeld 2.3B, blz. 221.

Onthoud: Cirkelsectoren gebruiken aandelen; histogrammen gebruiken grenzen; polygonen gebruiken middens.

6 Van tabel naar lijnen en staven

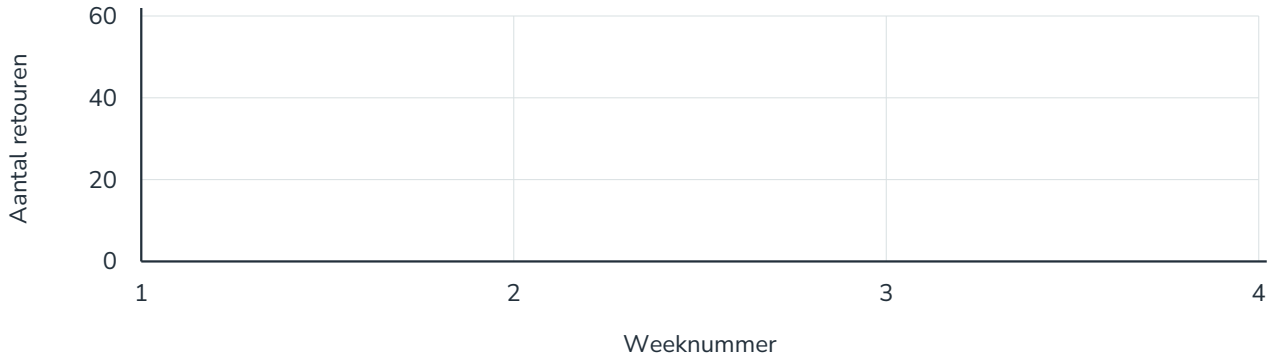
Een balie telt retouren per week en onderzoekt afzonderlijk twintig wachttijden. De gegevens zijn samengesteld.

Tabel I: week	Retouren
1	32
2	48
3	40
4	56

a Benoem de assen voor een lijndiagram van tabel I en schrijf de vier meetpunten op.

6 Van tabel naar lijnen en staven (vervolg)

Retouren per week



b Teken het lijndiagram met de punten uit a. Verbind opeenvolgende weken.

Tabel II: wachttijd t (min)	Frequentie
$0 \le t < 10$	4
$10 \le t < 20$	9
$20 \le t < 30$	6
$30 \le t \le 40$	1

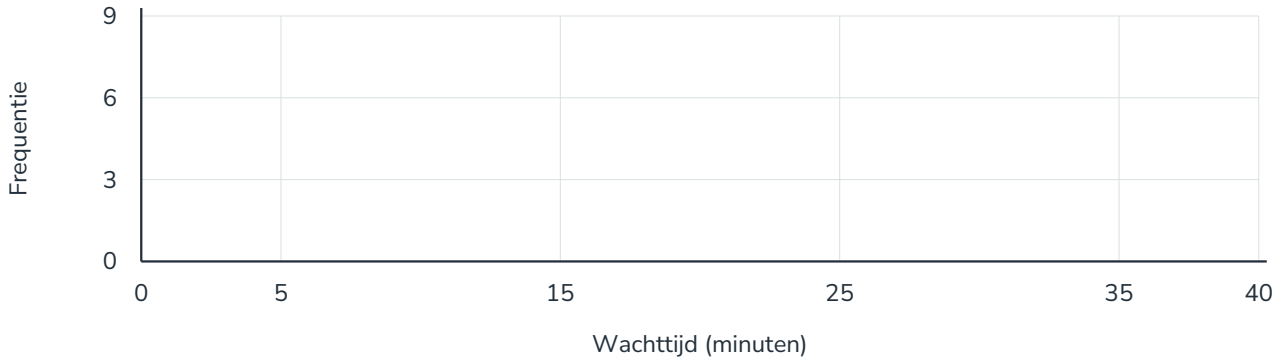
Wachttijden



c Teken bij tabel II een histogram. Gebruik de klassegrenzen en frequenties. Waarom sluiten de staven aan?

6 Van tabel naar lijnen en staven (vervolg)

Frequentiepolygoon wachttijd



d Bereken de klassenmiddens van tabel II bij c en teken het frequentiepolygoon. Waar hoort precies 20 minuten?

7 Twee beelden van dezelfde keuzes

Twintig bezoekers kiezen één route. De gegevens zijn samengesteld.

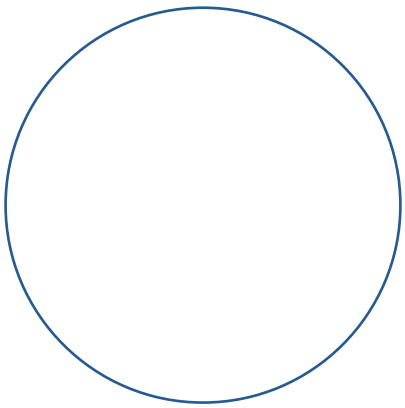
Route	Aantal
Bos	4
Water	8
Stad	8

Routekeuzes



a Teken een staafdiagram van de aantallen.

Aandelen routekeuzes

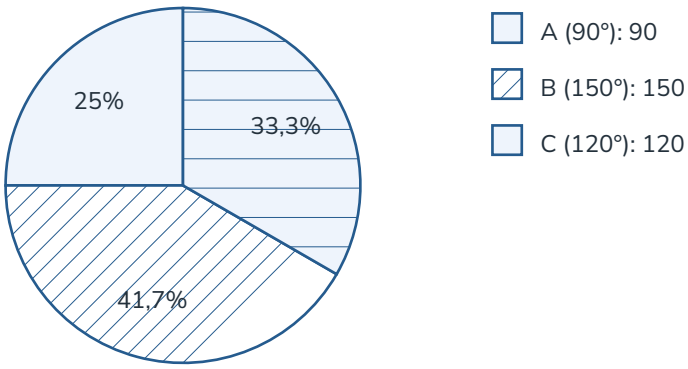


b Teken een cirkeldiagram. Geef bij elke sector het percentage en bereken de sectorhoeken.

8 Terug van cirkel naar aantallen

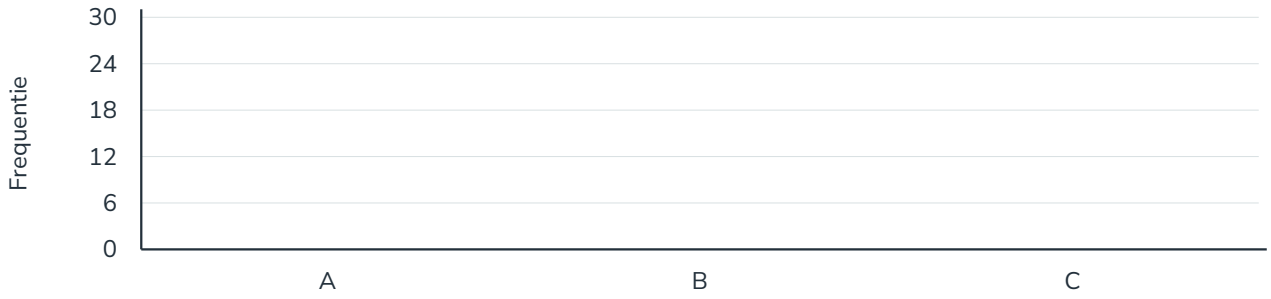
Van 72 deelnemers is één keuze geregistreerd. Het cirkeldiagram is opgebouwd uit de aangegeven sectorhoeken. De gegevens zijn samengesteld.

Sectorhoeken van de keuzes



a Bereken de aantallen bij A, B en C.

Keuzes van 72 deelnemers



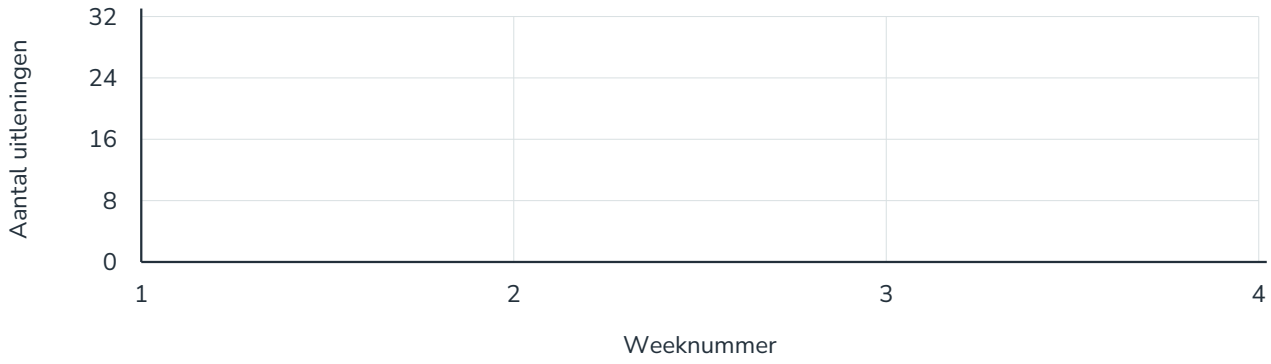
b Teken een staafdiagram van die aantallen. Controleer het totale aantal.

9 Ontwikkeling of verdeling

Een uitleenbalie verzamelt twee datasets. Tabel I telt uitleningen per week; tabel II groepeert twintig retourtijden. De gegevens zijn samengesteld.

Tabel I: week	Uitleningen
1	20
2	28
3	24
4	32

Uitleningen

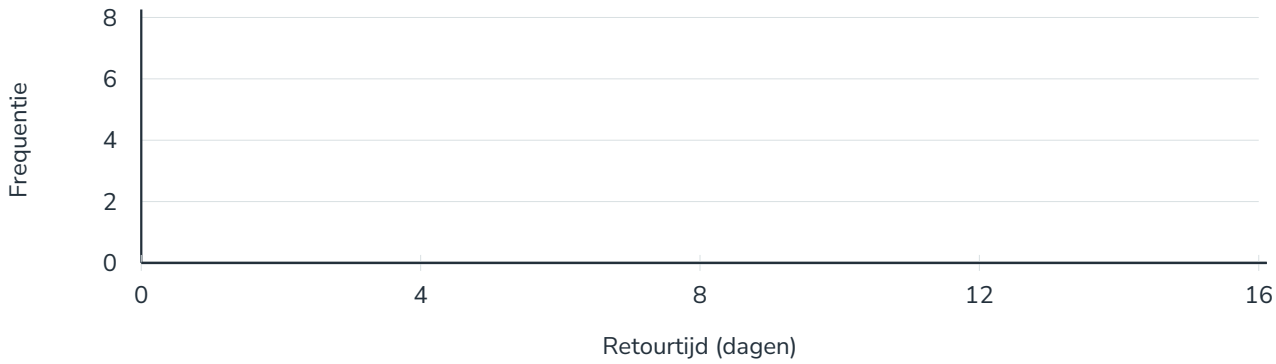


a Kies en teken een diagram dat de ontwikkeling in tabel I toont. Leg je keuze uit.

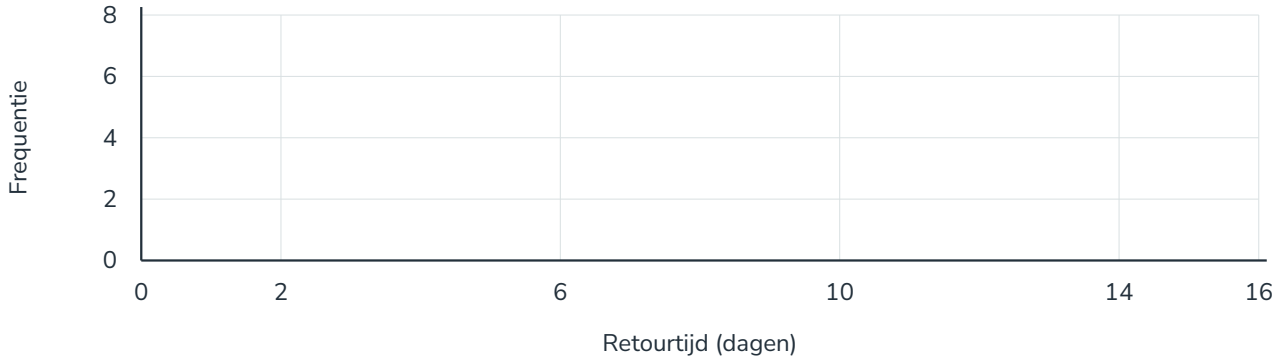
9 Ontwikkeling of verdeling (vervolg)

Tabel II: retourtijd t (dagen)	Frequentie
$0 \le t < 4$	3
$4 \le t < 8$	7
$8 \le t < 12$	8
$12 \le t \le 16$	2

Retourtijden



Frequentiepolygoon retourtijd

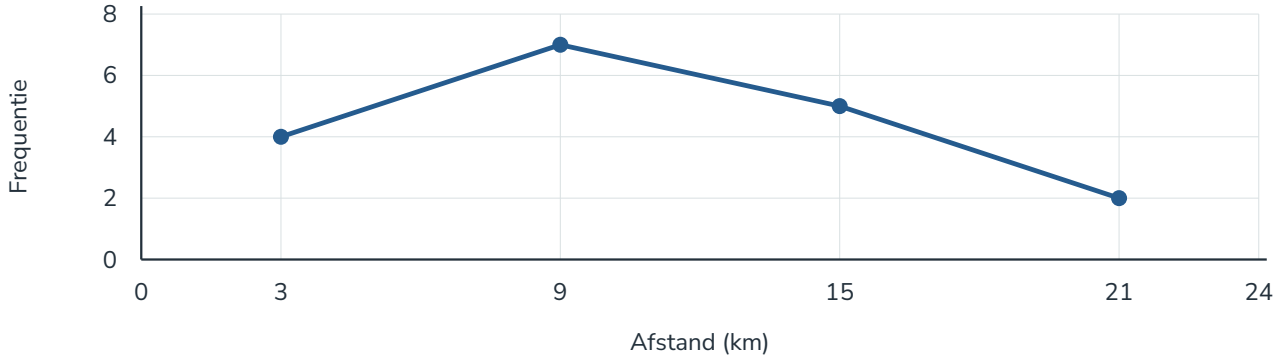


b Teken bij tabel II een histogram én een frequentiepolygoon. Welke klasse heeft de hoogste frequentie? Leg ook uit waar precies 8 dagen hoort.

10 Klassen uit een polygoon halen

Een frequentiepolygoon geeft afstanden van achttien ritten. Alle klassen zijn 6 km breed; de eerste begint bij 0 km en de laatste bovengrens hoort bij de laatste klasse. De gegevens zijn samengesteld.

Afstanden van ritten



a Maak de frequentietabel met volledige klassegrenzen.

b Welk percentage van de ritten is ten minste 12 km? Rond af op één decimaal.

Ordenen en gegevens koppelen

Leerdoel

Ik kan cumulatieve voorstellingen, steelbladdiagrammen en dotplots uit brondata opstellen en omzetten, en bij voldoende gekoppelde brondata een kruistabel of puntenwolk maken en lezen.

Theorie

Ordenen en gegevens koppelen

Een cumulatieve frequentie telt hoeveel waarnemingen je tot een grens hebt meegeteld. Tel de gewone frequenties op in hun natuurlijke volgorde. Voor een cumulatief percentage deel je ieder oplopend totaal door de hele dataset en vermenigvuldig je met 100%. Het laatste cumulatieve aantal is n; het laatste percentage is 100%.

Teken bij gegroepeerde data ieder cumulatief totaal op de bovengrens van de klasse en voeg het beginpunt op de eerste ondergrens met hoogte 0 toe. Verbind de punten tot een cumulatieve curve. Bij klassen $a \le x < b$ telt de grens b de waarden $x < b$; bij een inclusieve laatste bovengrens telt ook die eindwaarde mee. Klassegrenzen leveren exacte cumulatieve aantallen. Een rechte lijn tussen grenzen geeft slechts een schatting: de afzonderlijke waarden binnen de klasse zijn onbekend.

Zie voorbeeld 2.3C, blz. 223.

Onthoud: Een cumulatieve curve kan niet dalen.

11 Oplopende aantallen

Twintig wachttijden zijn gegroepeerd. De gegevens zijn samengesteld.

Wachttijd t (min)	Frequentie
$0 \le t < 5$	2
$5 \le t < 10$	6
$10 \le t < 15$	8
$15 \le t \le 20$	4

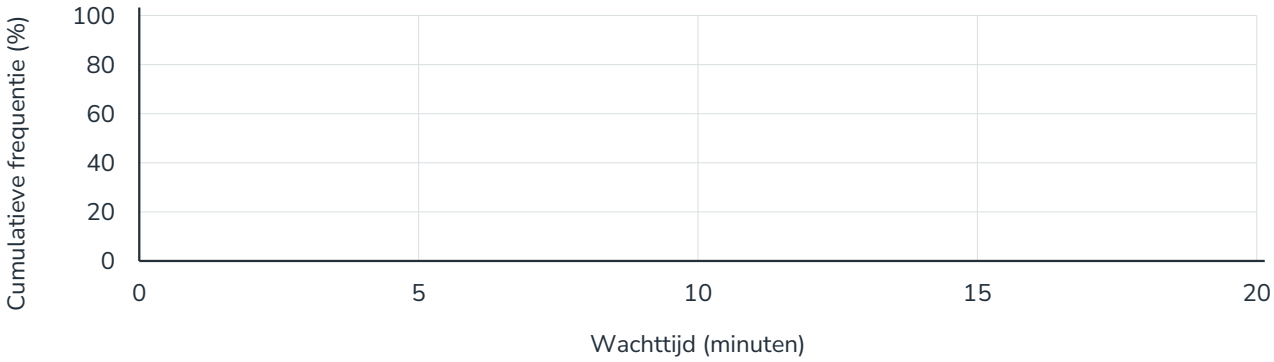
a Maak de cumulatieve tabel. Tel de frequenties van links naar rechts in de klassevolgorde op en reken elk totaal om naar een percentage van 20.

11 Oplopende aantallen (vervolg)

Twintig wachttijden zijn gegroepeerd. De gegevens zijn samengesteld.

Wachttijd t (min)	Frequentie
$0 \leq t < 5$	2
$5 \leq t < 10$	6
$10 \leq t < 15$	8
$15 \leq t \leq 20$	4

Cumulatieve wachttijd



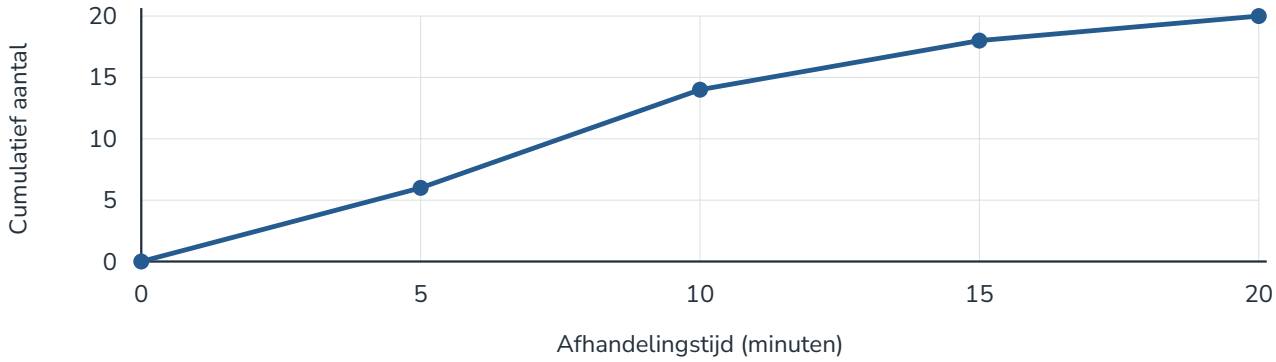
b Teken de cumulatieve procentcurve. Zet ieder percentage bij de bovengrens en begin bij (0, 0).

c Hoeveel procent wacht ten minste 10 minuten? Geef het antwoord uit de tabel of de curve.

12 Verschillen tussen cumulatieve aantallen

De cumulatieve curve hoort bij twintig afhandelingstijden. De eerste tijdklasse begint bij 0 minuten, alle klassen zijn 5 minuten breed en de laatste grens 20 is inclusief. De gegevens zijn samengesteld.

Cumulatieve afhandelingstijden



a Lees het cumulatieve aantal bij 10 minuten af. Welke tijden telt dit aantal mee?

b Maak de gewone frequentietabel. Neem de verschillen tussen opeenvolgende cumulatieve aantallen.

c De rechte lijnstukken geven een schatting tussen klassegrenzen. Schat het percentage korter dan 12,5 minuten. Is het exacte aantal uit deze curve bekend?

Theorie

Losse waarden bewaren

Een steelbladdiagram splitst elk getal in een stam en een blad. Bij gehele getallen kun je de tientallen als stam en de eenheden als bladeren nemen. Sorteer de bladeren en herhaal een blad voor iedere herhaalde waarde. Geef altijd een sleutel met eenheid. Een dotplot zet één stip per waarneming boven de getallenlijn. Gelijke waarden krijgen stippen boven elkaar. Beide voorstellingen bewaren de afzonderlijke waarden; een klassentabel verbergt juist de verdeling binnen elke klasse.

Onthoud: Eén blad of stip staat voor één waarneming.

13 Losse waarden bewaren

Van acht proefstroken zijn de lengtes in hele centimeters geregistreerd: 12, 14, 14, 17, 21, 21, 23 en 28. De gegevens zijn samengesteld; lengte blijft een continue variabele.

Stam	Bladeren
1	
2	

a Maak een steelbladdiagram. Geef de sleutel.

Lengtes van proefstroken



b Teken een dotplot van de acht lengtes.

c Maak een cumulatieve procenttabel bij de grenzen 20 en 30 cm, met klassen $10 \le l < 20$ en $20 \le l \le 30$. Welk aandeel is ten minste 20 cm?

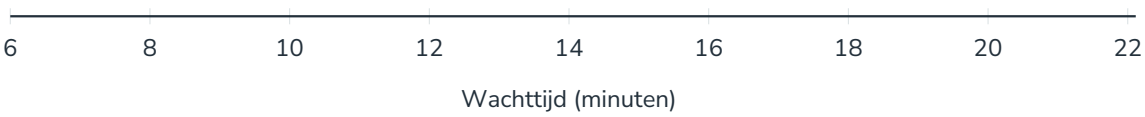
14 Van bladeren naar stippen

Het steelbladdiagram geeft acht wachttijden. Sleutel: 1 | 2 betekent 12 minuten. Deze gegevens zijn samengesteld.

Stam	Bladeren
0	6 8 8
1	0 2 2 9
2	1

a Schrijf alle wachttijden op en bepaal het verschil tussen de langste en de kortste.

Wachttijden

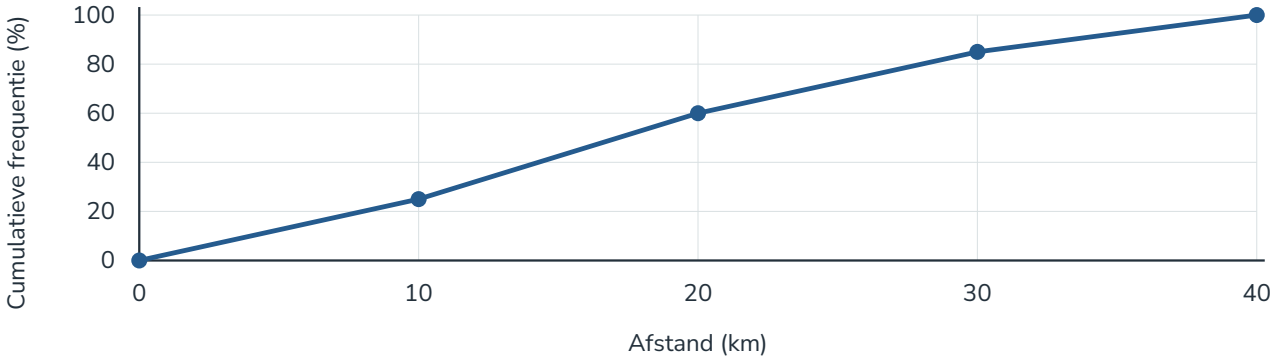


b Zet het steelbladdiagram om in een dotplot. Welk percentage wacht ten minste 12 minuten?

15 Terugrekenen uit percentages

De cumulatieve procentcurve beschrijft de afstanden van veertig ritten. De klassen zijn $0 \le d < 10$, $10 \le d < 20$, $20 \le d < 30$ en $30 \le d \le 40$ km. De gegevens zijn samengesteld.

Cumulatieve ritafstand



a Maak de gewone frequentietabel.

b Beoordeel de twee uitspraken: 85% rijdt minder dan 30 km; 70% rijdt minder dan 25 km. Welke uitspraak volgt exact uit de gegevens? Licht toe.

Theorie

Gekoppelde gegevens

Bij gekoppelde gegevens horen twee waarnemingen bij dezelfde persoon of gebeurtenis. Een kruistabel telt combinaties van categorieën. Tel iedere bronrij één keer; de rij- en kolomtotalen moeten samen dezelfde omvang geven. Voor percentages binnen een groep deel je door het totaal van die groep. Een puntenwolk gebruikt twee kwantitatieve variabelen: teken voor ieder bronpaar één punt (x, y), zonder verbindingslijnen. Beschrijf of grotere x doorgaans samengaat met grotere of kleinere y. Een zichtbaar verband bewijst geen oorzaak.

Onthoud: Houd twee gegevens van dezelfde waarneming bij elkaar.

16 Combinaties tellen

Van tien leerlingen zijn vervoermiddel en op tijd komen bij dezelfde aankomst geregistreerd. Deze gegevens zijn samengesteld.

Leerling	Vervoermiddel	Op tijd
1	Fiets	Ja
2	Fiets	Ja
3	Bus	Nee
4	Fiets	Nee
5	Bus	Ja
6	Bus	Nee
7	Fiets	Ja
8	Bus	Ja
9	Fiets	Ja
10	Bus	Nee

a Maak een kruistabel met vervoermiddel als rijen en op tijd als kolommen. Voeg rij- en kolomtotalen toe.

b Bereken per vervoermiddel het percentage dat op tijd is. Gebruik bij beide percentages de juiste basis.

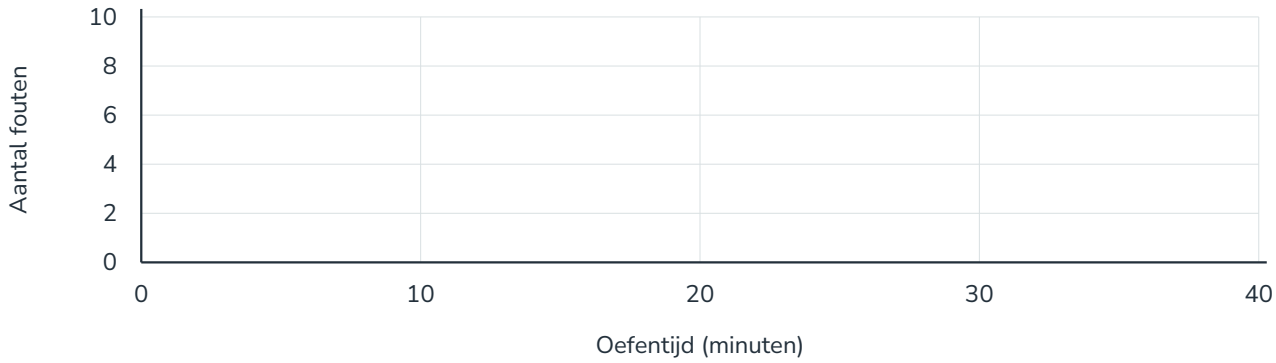
c Kun je uit deze tabel besluiten dat fietsen ervoor zorgt dat je op tijd komt? Leg uit.

17 Een patroon in paren

Van zes leerlingen zijn oefentijd en aantal fouten op dezelfde taak geregistreerd. De gegevens zijn samengesteld.

Leerling	Oefentijd (min)	Aantal fouten
1	10	8
2	10	6
3	20	7
4	20	5
5	30	4
6	30	3

Oefentijd en fouten



a Teken een puntenwolk met oefentijd horizontaal en aantal fouten verticaal.

b Beschrijf het patroon en de verschillen bij dezelfde oefentijd. Is een oorzakelijk gevolg bewezen?

18 Losse totalen zijn niet genoeg

Een overzicht geeft alleen twee losse verdelingen van dezelfde tien leerlingen. De gegevens zijn samengesteld.

Vervoermiddel	Aantal
Fiets	6
Bus	4

Op tijd	Aantal
Ja	7
Nee	3

a Kun je hiermee bepalen hoeveel fietsers op tijd zijn? Leg uit welke informatie ontbreekt.

b Geef twee verschillende kruistabellen die allebei bij deze totalen passen.

Een presentatie kritisch lezen

Leerdoel

Ik kan een datapresentatie beoordelen op schaal, weggelaten informatie en vergelijkbaarheid.

Theorie

Een presentatie kritisch lezen

Lees vóór je conclusies trekt de titel, assen, eenheden, schaal en periode. Een staafdiagram hoort bij 0 te beginnen als de staafhoogten aantallen of aandelen vergelijken. Een ingekorte as kan een klein verschil veel groter laten lijken. Bereken veranderingen met de werkelijke waarden; deel de toename voor procentuele groei door de beginwaarde. Een zichtbare hoogteverhouding kan daarvan afwijken.

Controleer ook wat ontbreekt. Zijn alle perioden getoond? Zijn groepen onder dezelfde omstandigheden gemeten? Vergelijk bij verschillende groepsgroottes percentages met het eigen groepstotaal als basis. Maandtotalen kunnen verschillen doordat het aantal meet- of openingsdagen verschilt; gebruik dan een passende hoeveelheid per dag. Geef duidelijk aan welke basis je kiest. Een figuur kan een verschil of verband tonen, maar bewijst daarmee geen oorzaak.

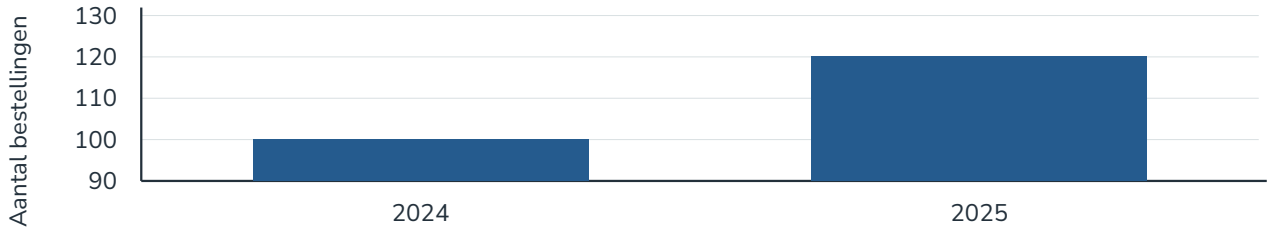
Zie voorbeeld 2.3D, blz. 229.

Onthoud: Controleer schaal, volledigheid en vergelijkingsbasis.

19 Een overdreven groei herstellen

Een advertentie toont deze twee jaartotalen en beweert: het aantal bestellingen is verdrievoudigd. De gegevens zijn samengesteld.

Bestellingen volgens de advertentie

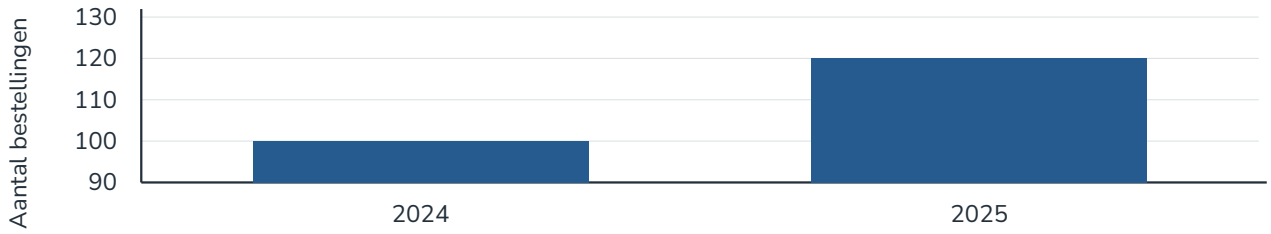


a Lees het begin van de verticale as. Bereken met de echte aantallen de procentuele groei.

19 Een overdreven groei herstellen (vervolg)

Een advertentie toont deze twee jaartotalen en beweert: het aantal bestellingen is verdrievoudigd. De gegevens zijn samengesteld.

Bestellingen volgens de advertentie



Bestellingen opnieuw tekenen



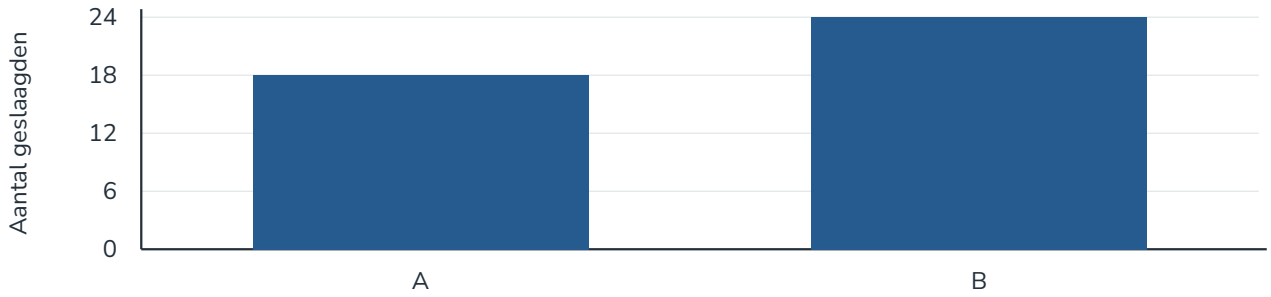
b Teken een verbeterd staafdiagram en vervang de bewering door een juiste zin.

20 Groepsgrootte maakt verschil

Twee groepen doen dezelfde vaardigheidstaak. De figuur toont alleen de aantallen die slagen. De gegevens zijn samengesteld.

Groep	Geslaagd	Deelnemers
A	18	30
B	24	60

Aantallen geslaagden



a Een verslag zegt: groep B presteert beter, want er slagen meer deelnemers. Beoordeel deze redenering met een berekening.

b Kies een voorstelling waarmee je de groepsresultaten eerlijk vergelijkt en teken die.

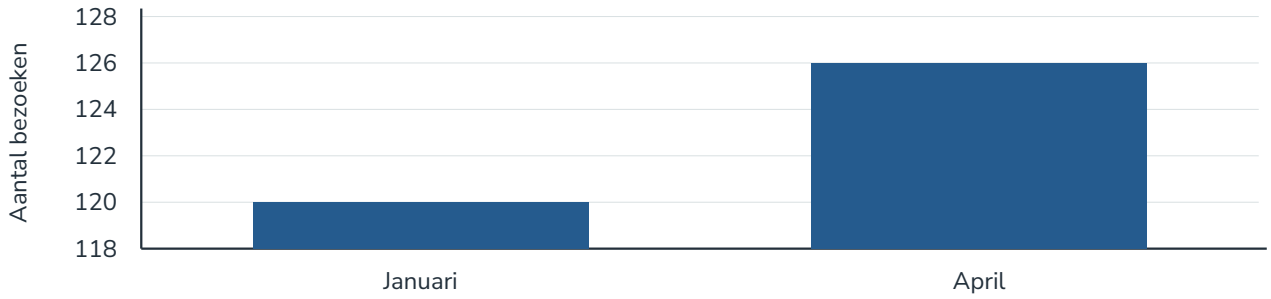
c Noem één gegeven over de meting dat je nog wilt weten voordat je de groepen inhoudelijk vergelijkt.

21 Wat laat een advertentie weg?

Een bibliotheek toont alleen januari en april en schrijft: iedere maand sterke groei. De volledige tabel en advertentiegegevens zijn samengesteld.

Maand	Bezoeken	Openingsdagen
Januari	120	20
Februari	160	20
Maart	180	20
April	126	14

De advertentie: twee maanden



a Kies en teken een voorstelling die de volledige ontwikkeling van het aantal bezoeken toont. Benoem de reden voor je keuze.

b Beoordeel de bewering iedere maand sterke groei. Bereken ook de verandering van januari naar april in procenten.

c Gebruik nu het aantal openingsdagen. Vergelijk maart en april op bezoeken per openingsdag. Wat verandert er aan je oordeel over de daling?

22 Nieuwe diagrammen maken

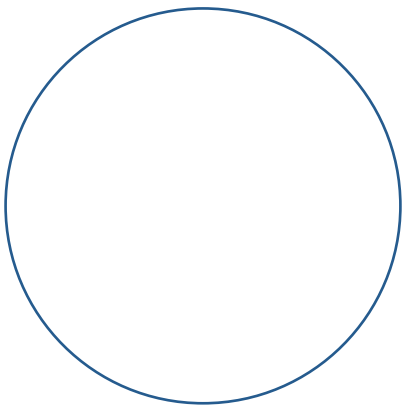
De drie tabellen komen uit drie afzonderlijke korte onderzoeken. De gegevens zijn samengesteld.

Tabel I: zaalkeuze	Aantal
A	6
B	9
C	3

Tabel II: taaktijd t (min)	Frequentie
$0 \le t < 10$	4
$10 \le t < 20$	8
$20 \le t \le 30$	6

Tabel III: week	Afspraken
1	15
2	12
3	18

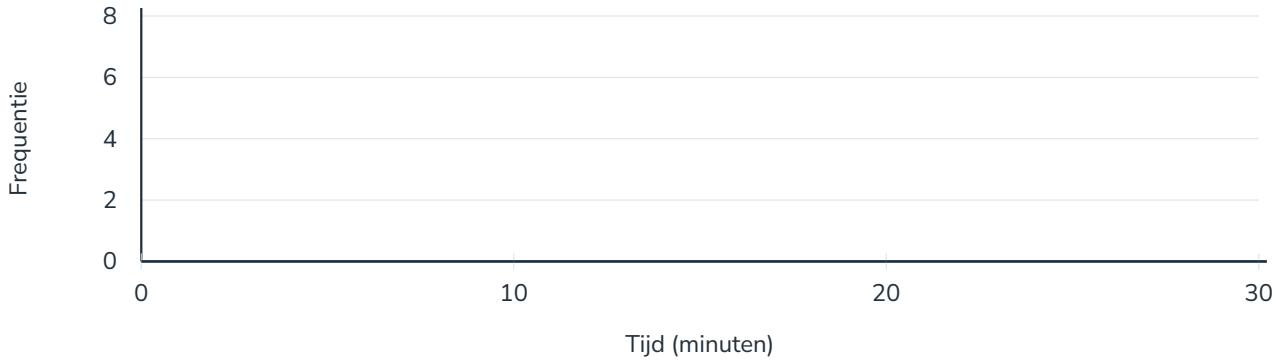
Aandelen zaalkeuzes



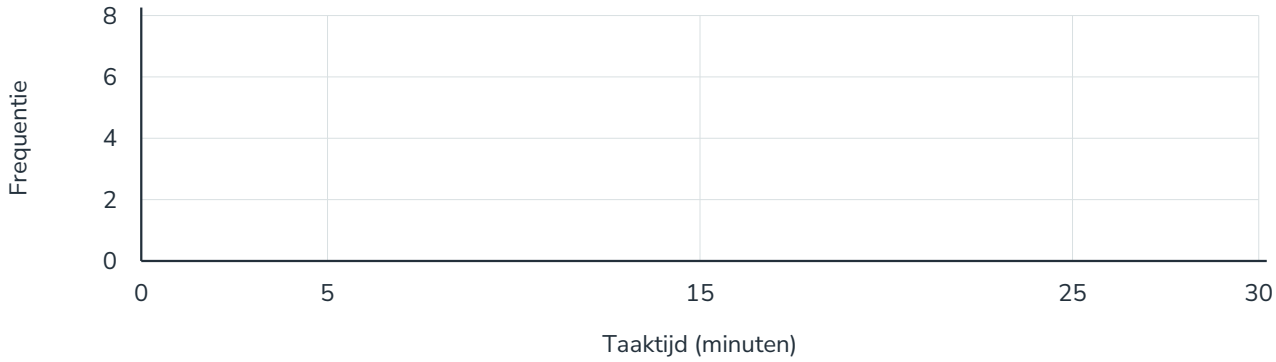
a Welk diagram zou je kiezen om de drie aantallen uit tabel I direct te vergelijken? Licht toe. Teken daarnaast een cirkeldiagram van de aandelen en bereken de sectorhoeken.

22 Nieuwe diagrammen maken (vervolg)

Taaktijden

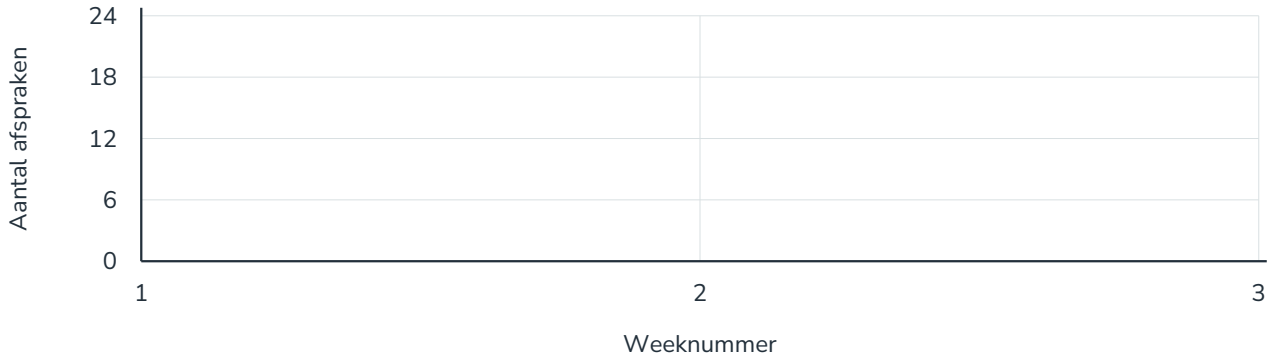


Frequentiepolygoon taaktijd



b Teken bij tabel II een histogram en een frequentiepolygoon.

Ontwikkeling afspraken



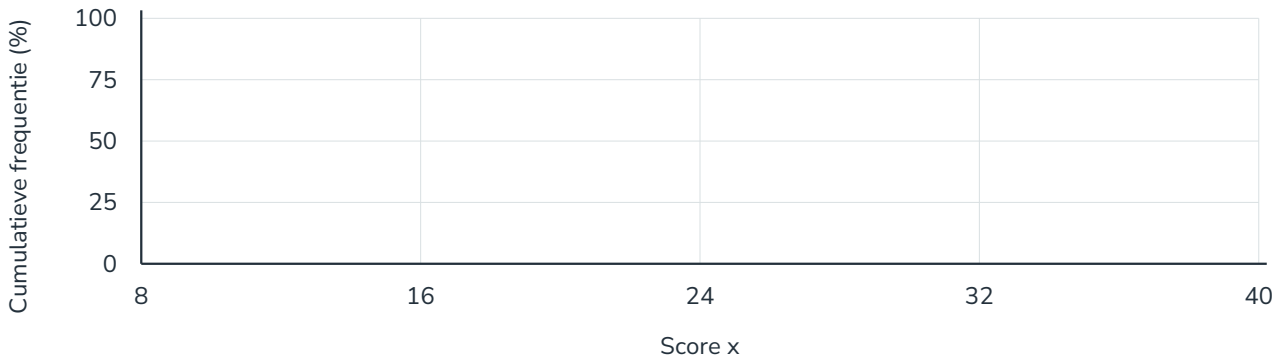
c Kies en teken een diagram dat de ontwikkeling in tabel III toont.

23 Dezelfde bron, verschillende beelden

Acht leerlingen doen twee taken. Per leerling is het aantal goede antwoorden x op taak A (maximum 40) en y op taak B (maximum 10) geregistreerd. De gegevens zijn samengesteld.

Leerling	Score x	Score y
1	12	4
2	16	5
3	16	6
4	20	6
5	24	6
6	24	7
7	28	8
8	32	9

Cumulatieve score op taak A



a Teken voor x de cumulatieve procentcurve met klassen $8 \le x < 16$, $16 \le x < 24$, $24 \le x < 32$ en $32 \le x \le 40$. Welk percentage heeft $x < 24$?

Stam	Bladeren
1	
2	
3	

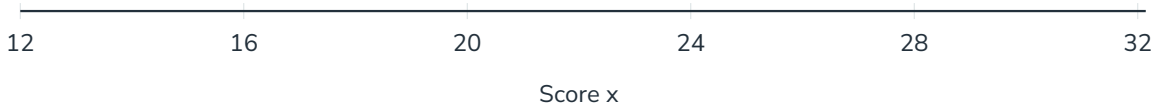
b Maak een steelbladdiagram van x en geef de sleutel.

23 Dezelfde bron, verschillende beelden (vervolg)

Acht leerlingen doen twee taken. Per leerling is het aantal goede antwoorden x op taak A (maximum 40) en y op taak B (maximum 10) geregistreerd. De gegevens zijn samengesteld.

Leerling	Score x	Score y
1	12	4
2	16	5
3	16	6
4	20	6
5	24	6
6	24	7
7	28	8
8	32	9

Scores op taak A



c Teken een dotplot van x.

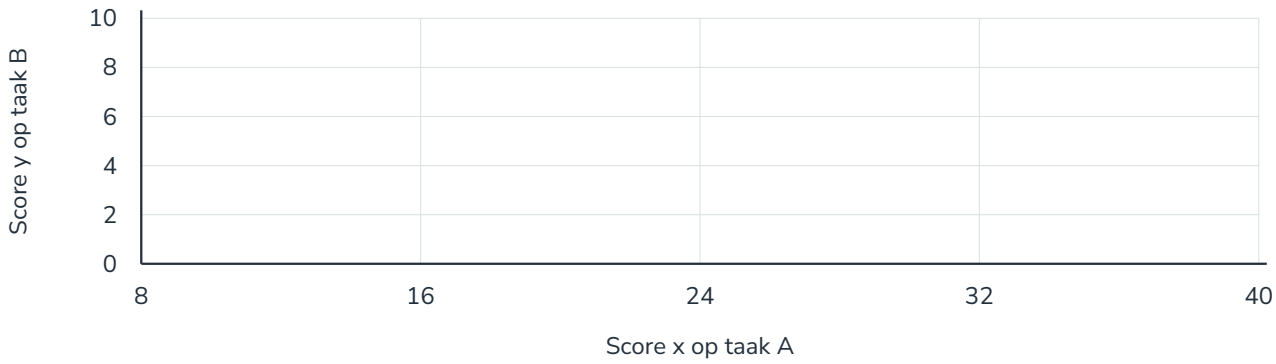
d Maak een kruistabel met rijen $x < 24$ en $x \geq 24$, en kolommen $y < 7$ en $y \geq 7$. Voeg totalen toe.

23 Dezelfde bron, verschillende beelden (vervolg)

Acht leerlingen doen twee taken. Per leerling is het aantal goede antwoorden x op taak A (maximum 40) en y op taak B (maximum 10) geregistreerd. De gegevens zijn samengesteld.

Leerling	Score x	Score y
1	12	4
2	16	5
3	16	6
4	20	6
5	24	6
6	24	7
7	28	8
8	32	9

Scores op beide taken

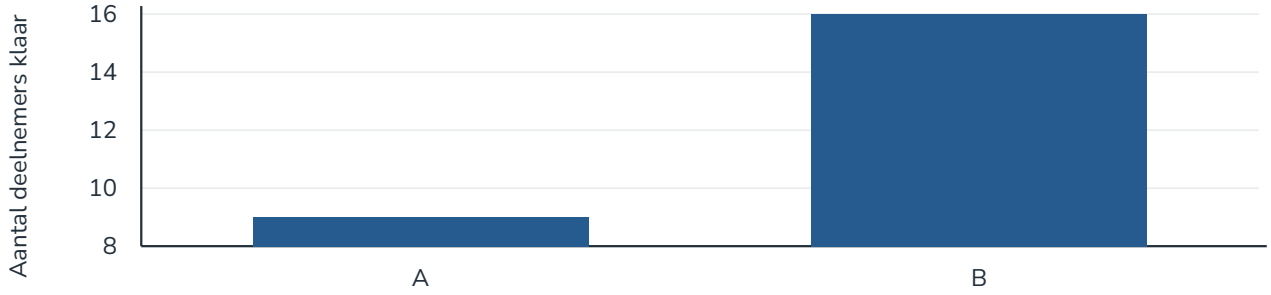


e Teken de puntenwolk van x en y. Leg de betekenis van (24, 7) uit en beschrijf het zichtbare patroon.

24 Een vergelijking beoordelen

De figuur toont de aantallen deelnemers die een taak afronden. In groep A zijn 9 van de 15 deelnemers klaar, in groep B 16 van de 40. De periode van meten is niet vermeld. Deze gegevens zijn samengesteld.

Resultaten



a Een verslag noemt B de succesvolste groep. Beoordeel dat oordeel met een passende berekening. Benoem ook één ontbrekend gegeven dat een inhoudelijke vergelijking kan veranderen.

b Beoordeel de verticale schaal van de gegeven figuur. Welk effect heeft deze schaal op de zichtbare hoogteverhouding?

c Teken een staafdiagram waarmee de resultaten van groepen met verschillende omvang vergelijkbaar zijn. Zorg dat het diagram zelfstandig leesbaar is.

2.4 Centrummaten berekenen en kiezen

Leerdoel

Ik kan gemiddelde, mediaan en modus bepalen uit losse gegevens of een frequentietabel.

Theorie

Gemiddelde, mediaan en modus

Een centrummaat vat het midden of een veelvoorkomende waarde van gegevens samen. Het gemiddelde is de som van alle getallen gedeeld door het aantal waarnemingen. Elke waarneming telt eenmaal mee. Bij een frequentietabel tel je daarom waarde × frequentie op en deel je door de som van de frequenties.

Voor de mediaan zet je de waarnemingen eerst in oplopende volgorde. Bij een oneven aantal neem je de middelste waarde. Bij een even aantal numerieke waarnemingen neem je het gemiddelde van de twee middelste waarden. Een frequentietabel helpt je de middelste plaatsen te vinden zonder alle waarden uit te schrijven. De modus is de waarde met de hoogste frequentie. Er kunnen meerdere modussen zijn. Als alle waarden eenmaal voorkomen, zeggen we hier dat er geen modus is. Zie voorbeeld 2.4A, blz. 231.

Onthoud: Bij een frequentietabel telt iedere waarde zo vaak mee als haar frequentie.

1 Ordenen en tellen

Gegeven zijn de getallen 9, 3, 6, 3, 4, 7 en 3.

a Zet de getallen in oplopende volgorde. Welk getal staat op de middelste plaats?

b Hoeveel getallen zijn er en wat is hun som?

2 Waarnemingen achter de tabel

De tabel geeft het aantal boeken dat tien leerlingen in een maand lazen. De gegevens zijn voor deze oefening samengesteld.

Aantal boeken	Frequentie
1	2
2	5
3	3

a Controleer met de frequenties hoeveel leerlingen in de tabel staan.

b Hoeveel boeken lazen deze leerlingen samen?

3 Drie maten voor één lijst

Zes bezoekers wachtten 9, 6, 7, 6, 12 en 8 minuten. De gegevens zijn voor deze oefening samengesteld.

a Bereken eerst de totale wachttijd en daarna het gemiddelde.

b Orden de wachttijden en bepaal de mediaan.

c Bepaal de modus. Leg uit welke frequentie je gebruikt.

4 Een ontbrekende score

Vijf scores zijn 6, 7, 7, 8 en een onbekende score. Hun gemiddelde is 7,4.

a Bepaal de onbekende score. Laat zien hoe je terugrekent.

b Iemand beweert dat de mediaan en de modus ook 7,4 zijn. Beoordeel de bewering.

5 Meer dan één modus

Vijftien leerlingen registreerden het aantal bezoeken aan een bibliotheek. De gegevens zijn voor deze oefening samengesteld.

Aantal bezoeken	Frequentie
0	3
1	5
2	5
3	2

a Bepaal het gemiddelde en de mediaan. Rond alleen het gemiddelde af op één decimaal.

b Bepaal alle modussen.

Rekenen met gewichten

Leerdoel

Ik kan een gewogen gemiddelde berekenen en uitleggen wanneer een gewoon gemiddelde ongeschikt is.

Theorie

Rekenen met gewichten

Een gewogen gemiddelde gebruik je als waarden niet even zwaar meetellen. Vermenigvuldig iedere waarde met haar gewicht, tel de producten op en deel door de som van de gewichten. Een gewicht kan een frequentie zijn, maar ook aangeven hoe zwaar een cijfer meetelt. Alle gewichten zijn niet-negatief en hun som moet positief zijn.

Bij het combineren van groepsgemiddelden zijn de groepsgroottes de gewichten. Groepsgrootte × groepsgemiddelde geeft de som in die groep. Tel de groepssommen op en deel door het totale aantal waarnemingen. Het gewone gemiddelde van twee groepsgemiddelden past alleen als beide groepen even groot zijn, al kan toevallig dezelfde uitkomst ontstaan. Om een ontbrekende waarde te vinden kun je terugrekenen: de totale som is aantal × gemiddelde. Rond pas de einduitkomst af. Zie voorbeeld 2.4B, blz. 233.

Onthoud: Deel de gewogen som door de som van de gewichten.

6 Niet elk cijfer telt even zwaar

Een eindcijfer wordt berekend uit de drie onderdelen in de tabel.

Onderdeel	Cijfer	Gewicht
Toets	6,0	1
Practicum	8,0	2
Eindopdracht	7,0	3

a Hoe vaak telt ieder cijfer mee als je de gewichten als aantallen opvat? Hoeveel cijfers zijn dat samen?

b Bereken het eindcijfer op één decimaal. Leg uit waarom (6,0 + 8,0 + 7,0) : 3 hier geen passende berekening is.

7 Een groepsgemiddelde reconstrueren

Een groep van 8 deelnemers gaf gemiddeld € 11 uit. Een andere groep telt 12 deelnemers. Alle 20 deelnemers samen gaven gemiddeld € 14 uit. De gegevens zijn voor deze oefening samengesteld.

a Bereken het gemiddelde bedrag van de tweede groep.

b Waarom levert het gewone gemiddelde van de twee groepsgemiddelden niet € 14 op?

8 Toevallig dezelfde uitkomst

Voor een project zijn de cijfers 9, 5 en 7 behaald, met achtereenvolgens gewicht 1, 1 en 2.

a Bereken het gewone en het gewogen gemiddelde.

b Een leerling zegt: «Bij ongelijke gewichten krijg je altijd verschillende uitkomsten.» Beoordeel dit met deze gegevens.

9 Een gecombineerde wachttijd

Bij afhaalpunt A wachtten 15 bezoekers gemiddeld 18 minuten. Bij afhaalpunt B wachtten 5 bezoekers gemiddeld 30 minuten. De gegevens zijn voor deze oefening samengesteld.

a Een beheerder noemt 24 minuten als gemiddelde van alle 20 bezoekers. Controleer en verbeter zijn uitkomst.

b Later komen er 5 bezoekers bij die elk 17 minuten wachten. Is het gemiddelde van de 25 bezoekers dan hoogstens 20 minuten?

Kiezen bij de onderzoeksvraag

Leerdoel

Ik kan een centrummaat kiezen bij het gegevenstype en de onderzoeksvraag.

Theorie

Kiezen bij de onderzoeksvraag

Kies een centrummaat bij zowel het gegevenstype als de onderzoeksvraag. Bij nominale categorieën, zoals vervoerswijze, is de modus bruikbaar voor de meest voorkomende keuze. Gemiddelde en mediaan hebben daar geen betekenis. Bij ordinale categorieën bestaat een natuurlijke volgorde: je kunt de modus of de middelste categorie gebruiken. Zijn er twee verschillende centrale categorieën, vermeld die dan beide; neem geen rekenkundig gemiddelde van hun codes.

Bij numerieke gegevens kun je gemiddelde, mediaan en modus berekenen. Voor een totaal of een verdeling van een totaal over personen is het gemiddelde handig. Voor een gebruikelijke waarde bij een opvallend grote of kleine waarneming is de mediaan vaak geschikter, omdat de grootte van die ene waarde minder invloed heeft. De modus beantwoordt de vraag welke waarde het vaakst voorkomt. Geef bij je keuze een reden vanuit de onderzoeksvraag.

Onthoud: Een geschikte centrummaat past bij het gegevenstype én bij de vraag.

10 Een centrummaat bij categorieën

Twintig leerlingen kozen hun favoriete gerecht. De gegevens zijn voor deze oefening samengesteld.

Gerecht	Frequentie
Pasta	6
Soep	5
Sandwich	9

a Welke van gemiddelde, mediaan en modus is geschikt om het meest gekozen gerecht te beschrijven? Bepaal die maat.

b Een onderzoeker geeft pasta code 1, soep code 2 en sandwich code 3. Hij krijgt een gemiddelde code van 2,15. Waarom is dit geen gemiddelde voorkeur?

11 De vraag bepaalt de maat

Een stagecoördinator verzamelde drie soorten gegevens. De gegevens zijn voor deze oefening samengesteld. De beoordelingen staan van slecht naar uitstekend in natuurlijke volgorde.

Vervoerswijze	Frequentie
Bus	8
Fiets	12
Lopen	5

Reistijden in minuten
4; 5; 5; 6; 6; 7; 37

Beoordeling	Frequentie
Slecht	2
Matig	3
Goed	4
Uitstekend	2

a De coördinator wil de meest gebruikte vervoerswijze noemen. Kies een centrummaat en geef de uitkomst.

b Voor de begroting wil hij de totale reistijd van de zeven geregistreerde reizen berekenen met «aantal × centrummaat». Welke maat werkt daarvoor? Bereken de maat en de totale tijd.

c Hij wil de middelste beoordeling van de elf leerlingen beschrijven. Kies een maat en bepaal de beoordeling. Waarom gebruik je geen gemiddelde van categoriecodes?

12 Twee rijen vergelijken

Bij twee kassarijen werden wachttijden gemeten, in minuten. De gegevens zijn voor deze oefening samengesteld.

Rij	Wachttijden in minuten
A	3; 4; 4; 5; 5; 6; 29
B	6; 7; 7; 8; 8; 9; 11

a Welke centrummaat geeft bij deze gegevens het bruikbaarste beeld van de gebruikelijke wachttijd: gemiddelde of mediaan? Bereken je gekozen maat voor beide rijen en vergelijk.

b De manager onderzoekt juist de totale wachttijd van de zeven gemeten bezoekers per rij. Welke van de twee maten gebruik je daarvoor? Verschillen de totalen?

13 Kledingmaten

Een winkel registreerde de maten van twintig verkochte shirts. De gegevens zijn voor deze oefening samengesteld. De maten staan in natuurlijke volgorde.

Maat	Frequentie
XS	2
S	4
M	8
L	5
XL	1

a Bepaal de modus en de mediaan van de verkochte maten.

b De winkel wil weten welke maat hij het meest moet aanvullen. Kies tussen modus en mediaan en leg je keuze uit.

Een uitschieter of een bewerking

Leerdoel

Ik kan de invloed van een uitschieter of een bewerking van alle waarnemingen op centrummaten voorspellen.

Theorie

Een uitschieter of een bewerking

Een uitschieter ligt opvallend ver van de overige waarnemingen. Een grote uitschieter kan het gemiddelde sterk omhoog trekken; een kleine juist omlaag. De mediaan gebruikt de centrale plaatsen en is vaak minder gevoelig. Dat betekent niet dat hij nooit verandert. Als je een waarneming toevoegt of verwijdert, kunnen de centrale plaatsen verschuiven. Ook de modus kan veranderen wanneer frequenties veranderen. Verwijder een waarneming alleen met een inhoudelijke reden, zoals een vastgestelde registratiefout.

Pas je op alle numerieke waarnemingen dezelfde bewerking $y = ax + b$ toe, met $a > 0$, dan veranderen gemiddelde en mediaan ook volgens die bewerking. Elke modus wordt eveneens $ax + b$. De volgorde blijft behouden en gelijke waarden blijven gelijk. Verander je slechts één waarneming, dan gelden deze regels voor alle waarden niet. Zie voorbeeld 2.4B, blz. 233.

Onthoud: Een bewerking van alle waarden is iets anders dan één waarde vervangen.

14 Eén waarde of alle waarden

De oorspronkelijke gegevens zijn 4, 4, 5, 7, 8 en 8.

a Eén van de twee waarden 8 wordt vervangen door 26. Voorspel wat er met gemiddelde, mediaan en modussen gebeurt. Onderbouw met de benodigde berekeningen.

b Ga weer uit van de oorspronkelijke gegevens. Bij elke waarde wordt 3 opgeteld, waarna de uitkomst wordt verdubbeld. Voorspel de drie centrummaten zonder alle nieuwe waarden uit te schrijven.

15 Een waarde verwijderen

Vijf bezorgtijden zijn 6, 7, 8, 9 en 40 minuten. Het tijdstip bij de bezorging van 40 minuten bleek verkeerd geregistreerd; die waarneming wordt uit deze berekening verwijderd. De gegevens zijn voor deze oefening samengesteld.

a Bereken gemiddelde en mediaan vóór en na het verwijderen.

b Waarom klopt de uitspraak «Een uitschieter verwijderen verandert de mediaan nooit» niet?

16 Welke redenering gaat mis?

De oorspronkelijke gegevens zijn 2, 4, 4, 6 en 9. Ze hebben gemiddelde 5, mediaan 4 en modus 4.

a 9 wordt vervangen door 29. Noor zegt: «De som stijgt met 20, dus alle drie de centrummaten stijgen met 4.» Verbeter haar redenering en geef de drie nieuwe maten.

b Ga weer uit van de oorspronkelijke gegevens. Elke waarde wordt omgezet met $y = 2x + 3$. Amir zegt: «Het gemiddelde wordt 13; mediaan en modus blijven 4, want de volgorde blijft hetzelfde.» Verbeter zijn conclusie.

17 Ritten in een week

De tabel geeft het aantal fietsritten van tien deelnemers in één week. Een tweede groep van 15 deelnemers maakte gemiddeld 2,6 ritten. De gegevens zijn voor deze oefening samengesteld.

Aantal ritten	Frequentie
0	2
1	5
2	3

a Bepaal het gemiddelde, de mediaan en de modus voor de tien deelnemers uit de tabel.

b Bepaal het gemiddelde voor alle 25 deelnemers samen. Een deelnemer stelt 1,85 voor. Beoordeel die uitkomst.

18 Reistijden beoordelen

Zeven reistijden naar school zijn 18, 19, 20, 20, 21, 22 en 90 minuten. De gegevens zijn voor deze oefening samengesteld.

a Je wilt een gebruikelijke reistijd voor deze groep beschrijven. Kies tussen gemiddelde en mediaan, bereken beide en onderbouw je keuze.

b Elke oorspronkelijke reistijd wordt omgezet met $y = 1{,}5(x + 3)$. Bepaal de nieuwe gemiddelde reistijd, mediaan en modus zonder alle nieuwe tijden uit te schrijven.

c Ga weer uit van de oorspronkelijke tijden. De tijd van 90 minuten blijkt een registratiefout en wordt verbeterd in 30 minuten. Bereken het nieuwe gemiddelde op één decimaal en bepaal de nieuwe mediaan.

2.5 Spreiding en boxplots

Leerdoel

Ik kan spreidingsbreedte, kwartielen en interkwartielafstand bepalen uit geordende data.

Theorie

Kwartielen en spreiding

Spreiding beschrijft hoe ver waarden uit elkaar liggen. Orden eerst de gegevens van klein naar groot. Het minimum is de kleinste waarde en het maximum de grootste. De spreidingsbreedte is maximum − minimum. De mediaan is de middelste waarde; bij een even aantal neem je het gemiddelde van de twee middelste waarden.

Verdeel de geordende reeks in een onderste en een bovenste helft. Bij een oneven aantal laat je de centrale mediaan buiten beide helften. Het eerste kwartiel Q1 is de mediaan van de onderste helft; het derde kwartiel Q3 is de mediaan van de bovenste helft. De interkwartielafstand is Q3 − Q1 en beschrijft de spreiding van de middelste helft. Beide spreidingsmaten hebben dezelfde eenheid als de gegevens. Gebruik hier deze handconventie; kwartieluitvoer van ICT kan een andere conventie volgen.

Zie voorbeeld 2.5A, blz. 235.

Onthoud: Orden eerst; laat bij een oneven aantal de centrale mediaan buiten de twee helften.

1 Ordenen en mediaan

Een oefenreeks bestaat uit 11, 7, 9, 13, 8, 10 en 12.

a Orden de waarden en bepaal de mediaan.

b Voeg de waarde 14 toe. Bepaal de nieuwe mediaan.

2 Waarden met frequenties

De tabel beschrijft een oefenreeks.

Waarde	Frequentie
2	2
3	4
5	2

a Hoeveel waarden bevat de reeks? Bepaal de mediaan.

3 Twee helften

Samengestelde oefengegevens: acht wachttijden in minuten zijn 2, 4, 5, 7, 8, 10, 11 en 13.

a Schrijf de onderste en de bovenste helft op. Bepaal Q1, de mediaan en Q3.

b Bereken de spreidingsbreedte en de interkwartielafstand.

4 Negen afwijkingen

Voor de oefening zijn negen positieve massaverschillen in gram samengesteld: 18, 12, 17, 13, 16, 14, 20, 15 en 19.

a Bereken de spreidingsbreedte en de interkwartielafstand. Laat zien welke kwartielen je gebruikt.

5 Een ontbrekende waarde

Een geordende oefenreeks is 2, 4, 6, 8, 10, a, 16, 18. De interkwartielafstand is 9.

a Bepaal a.

b Vervang a door 14. Wat worden de mediaan, spreidingsbreedte en interkwartielafstand?

Een boxplot gebruiken

Leerdoel

Ik kan een boxplot maken en waarden of informatie over aantallen uit een boxplot afleiden.

Theorie

Een boxplot gebruiken

Een boxplot laat vijf waarden zien: minimum, Q1, mediaan, Q3 en maximum. Teken op één getallenlijn een box van Q1 tot Q3. Zet een lijn in de box bij de mediaan. Verbind de box met de uiteinden bij minimum en maximum. In dit hoofdstuk lopen die uiteinden altijd tot minimum en maximum; er worden geen aparte uitschieterpunten getekend.

De kwartielen verdelen de geordende waarden ongeveer in vier even grote groepen. De lengte van een deel toont spreiding, geen aantal. Binnen een deel zijn waarden niet noodzakelijk gelijkmatig verdeeld. Bij herhaalde waarden kunnen meerdere waarnemingen op een kwartielgrens liggen. Je kunt daarom niet altijd exacte aantallen of percentages aflezen. Vergelijk boxplots op dezelfde schaal. De mediaan vergelijkt het centrum, de boxlengte de interkwartielafstand en de totale lengte de spreidingsbreedte. Afzonderlijke waarden en de standaardafwijking zijn meestal niet terug te vinden.

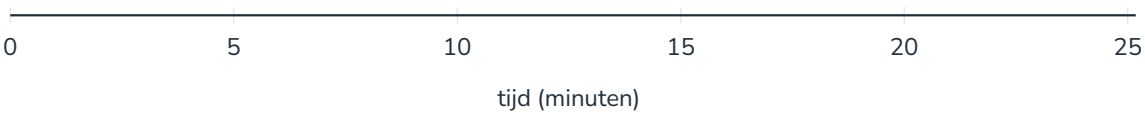
Zie voorbeeld 2.5A, blz. 235.

Onthoud: Een lange box bevat niet meer waarden; hij laat meer spreiding van de middelste helft zien.

6 Een boxplot tekenen

Samengestelde oefentijden (min): min = 3, Q1 = 7, mediaan = 10, Q3 = 16, max = 24.

Bezorgtijd



a Teken de boxplot op de as. Geef de vijf kenmerkende waarden aan.

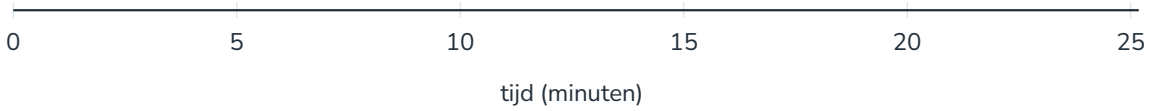
6 Een boxplot tekenen (vervolg)

b Welk interval hoort bij het laagste kwart van de geordende waarden? Bereken ook de interkwartielafstand.

7 Van lijst naar boxplot

Samengestelde oefengegevens: twaalf looptijden in minuten zijn 5, 6, 7, 8, 9, 10, 12, 14, 15, 17, 19 en 22.

Looptijd



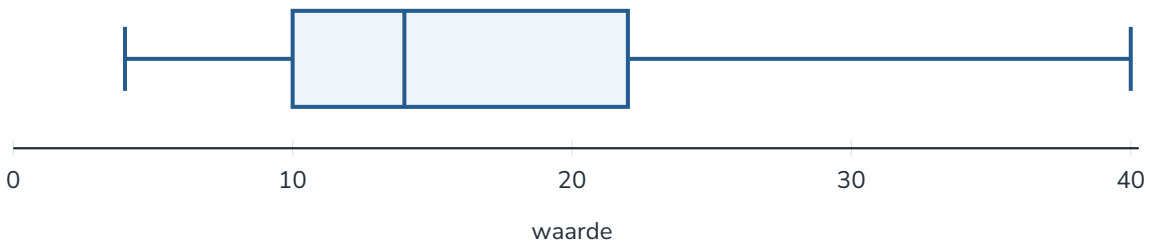
a Maak een boxplot van deze gegevens.

b Hoeveel van deze twaalf tijden zijn groter dan Q3? Controleer je antwoord in de lijst.

8 Wat weet je over aantallen?

Een boxplot hoort bij 24 samengestelde oefenwaarden. Alle waarden zijn verschillend. Er is volgens de handconventie gerekend.

Oefenwaarden



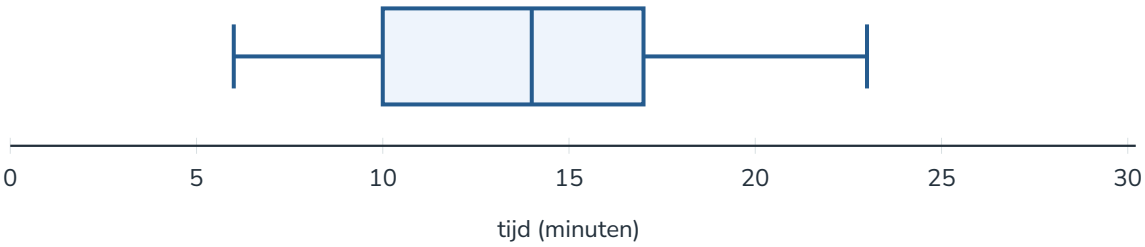
a Hoeveel waarden zijn kleiner dan 10? Hoeveel liggen tussen 10 en 22?

b Wat is het kleinste en het grootste mogelijke aantal waarden dat groter is dan 28? Licht toe.

9 Twee routes vergelijken

Voor de oefening zijn per route veertig reistijden samengesteld. Alle tijden binnen een route zijn verschillend. De boxplots gebruiken de handconventie.

Route A



Route B



a Welke route heeft de kortste mediane reistijd? Welke route heeft de kleinste spreiding van de middelste helft? Onderbouw met getallen.

b Kun je bepalen welke route vaker een reistijd van meer dan 20 minuten heeft? Leg uit.

Standaardafwijking met ICT

Leerdoel

Ik kan standaardafwijking met geschikte ICT bepalen en de keuze tussen standaardafwijking, interkwartielafstand en spreidingsbreedte beargumenteren.

Theorie

Standaardafwijking met ICT

De standaardafwijking beschrijft de spreiding van alle waarden rond het gemiddelde. Voor de beschreven dataset gebruiken we σx = √(som van (waarde − gemiddelde)² / n). Voer elke waarde in een gegevenslijst in. Herhalingen tellen mee; verwijder oude invoer en behandel lege plaatsen niet als nullen. Bij een frequentietabel voer je waarden en frequenties in aparte lijsten in en selecteer je de frequentielijst als frequentiebron. Open de statistiek voor één variabele en controleer n en het gemiddelde. Lees σx af, niet sx: sx deelt door n − 1.

Rond pas de einduitkomst af en vermeld de eenheid. Een grotere σx betekent meer spreiding rond het gemiddelde; σx is geen maximale afstand en geeft geen vast percentage binnen een interval. Kies spreidingsbreedte voor de uitersten, interkwartielafstand voor de middelste helft en σx als alle afwijkingen van het gemiddelde moeten meetellen.

Zie voorbeeld 2.5B, blz. 237.

Onthoud: Gebruik voor deze dataset σx met deler n en controleer of de GR het juiste aantal waarden gebruikt.

10 Eén variabele in de GR

Voor de oefening zijn zes wachttijden samengesteld: 4, 6, 6, 8, 10 en 14 minuten.

a Voer de waarden in één lijst in en open de statistiek voor één variabele. Controleer n = 6. Noteer het gemiddelde en de standaardafwijking σx. Rond σx af op twee decimalen.

b Controleer de uitkomst door de som van de gekwadrateerde afwijkingen van het gemiddelde te berekenen.

c Welke spreidingsmaat gebruik je voor de middelste helft van deze tijden? Welke gebruik je voor het verschil tussen de uiterste tijden? Leg beide keuzes uit.

11 Frequenties invoeren

De tabel bevat samengestelde oefengegevens over lengtes.

Lengte (cm)	Frequentie
12	1
14	3
16	3
18	1

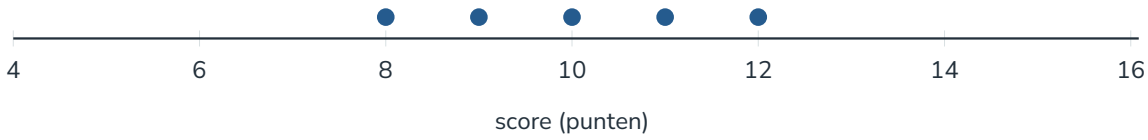
a Voer de waarden in een gegevenslijst en de frequenties in een tweede lijst in. Gebruik de tweede lijst als frequentiebron bij éénvariabele statistiek. Bepaal n, het gemiddelde en σx op twee decimalen.

b Zonder frequentielijst krijg je σx ≈ 2,24 cm. Verklaar het verschil.

12 Hetzelfde gemiddelde

De dotplots bevatten samengestelde oefenscores van twee groepen. Elk punt stelt één score voor.

Groep A



Groep B



a Bepaal met de GR voor beide groepen het gemiddelde en σx op één decimaal. Welke groep heeft de kleinste spreiding rond het gemiddelde?

b Geef vijf scores met gemiddelde 10 en standaardafwijking 0. Leg uit waarom jouw reeks voldoet.

c Je wilt de spreiding rond het gemiddelde vergelijken en alle scores meewegen. Kies tussen standaardafwijking, interkwartielafstand en spreidingsbreedte. Beargumenteer je keuze.

13 Een maat kiezen

Samengestelde oefengegevens: zeven reparaties duren 3, 4, 4, 5, 6, 6 en 24 minuten.

a Bereken σx op één decimaal en de interkwartielafstand.

b Je wilt de spreiding van de middelste helft beschrijven. Kies tussen σx, interkwartielafstand en spreidingsbreedte. Welke maat gebruik je om het verschil tussen de uiterste tijden te beschrijven?

14 Eén getal is niet genoeg

De tabel bevat twee samengestelde oefenreeksen met gemeten massaverschillen.

Reeks	Massaverschillen (g)
A	30, 39, 39, 40, 40, 41, 41, 50
B	30, 34, 38, 40, 40, 42, 46, 50

a Vergelijk de reeksen met spreidingsbreedte en σx. Rond σx af op twee decimalen.

b Een onderzoeker wil vooral weten hoe dicht de middelste helft bij elkaar ligt. Bepaal daarvoor een passende maat voor beide reeksen.

Uitschieters en bewerkingen

Leerdoel

Ik kan uitleggen wat een uitschieter of een bewerking van alle waarden doet met spreidingsmaten.

Theorie

Uitschieters en bewerkingen

Een waarde ver buiten de overige waarden noemen we hier een uitschieter. Er wordt geen vaste grensregel gebruikt. Als een uiterste waarde verder van de rest komt te liggen, veranderen de spreidingsbreedte en meestal de standaardafwijking sterk. De interkwartielafstand is vaak minder gevoelig, maar blijft niet altijd gelijk. Controleer de kwartielen: bij kleine reeksen kan het maximum meetellen in Q3. Toevoegen van een waarde kan bovendien de plaatsen van mediaan en kwartielen veranderen.

Bij $y = ax + b$ met $a > 0$ blijven de waarden in dezelfde volgorde. Minimum, kwartielen, mediaan en maximum worden elk vermenigvuldigd met a en daarna met b verhoogd. Verschillen tussen waarden worden alleen vermenigvuldigd met a. Daarom worden spreidingsbreedte, interkwartielafstand en standaardafwijking a keer zo groot. Alleen b optellen verschuift de gegevens zonder de spreiding te veranderen. Dit geldt ook bij het omrekenen van eenheden.

Zie voorbeeld 2.5B, blz. 237.

Onthoud: Bij $y = ax + b$ met $a > 0$ worden alle spreidingsmaten met a vermenigvuldigd.

15 **Eén extreme waarde**

Een oefenreeks tijden is 2, 3, 4, 5, 6, 7, 8 minuten. Vervang de laatste tijd door 20 minuten.

a Vergelijk de oude en de nieuwe spreidingsbreedte en interkwartielafstand. Bepaal daarvoor bij beide reeksen Q1 en Q3.

b Bepaal beide standaardafwijkingen op één decimaal. Verklaar waarom de nieuwe standaardafwijking groter is.

16 Verandert de box altijd?

Een oefenreeks is 2, 4, 6, 8. Vervang 8 door 28.

a Bereken bij beide reeksen Q1, Q3 en de interkwartielafstand.

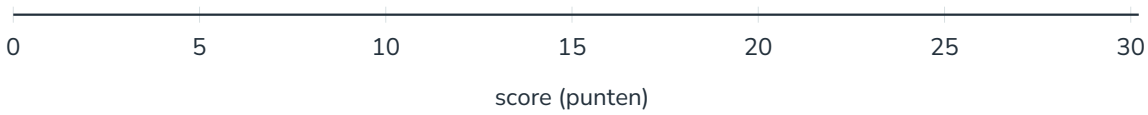
b Beoordeel de uitspraak: «Het vervangen van het maximum door een uitschieter verandert de interkwartielafstand nooit.»

17 Een nieuwe schaal

Een oefenreeks scores heeft minimum 2, Q1 = 4,5, mediaan 7,5, Q3 = 10,5 en maximum 13 punten. De standaardafwijking is exact 3,5 punten. Elke score x wordt vervangen door y = 2x + 3.

a Bepaal zonder de volledige gegevenslijst de nieuwe mediaan, spreidingsbreedte, interkwartielafstand en standaardafwijking.

Nieuwe scores



b Teken de boxplot van de nieuwe scores.

c Bij een andere omzetting y = ax + b met a > 0 worden de mediaan 20 punten en de standaardafwijking 10,5 punten. Bepaal a en b.

18 Van centimeter naar meter

Samengestelde oefengegevens: vijf lengtes zijn 160, 162, 165, 168, 170 cm. De standaardafwijking is afgerond 3,7 cm.

a Geef de mediaan, interkwartielafstand en de gegeven afgeronde standaardafwijking in meters.

b Er wordt bij elke oorspronkelijke lengte 2 cm opgeteld en daarna naar meters omgerekend. Wat veranderen de mediaan en interkwartielafstand ten opzichte van a?

19 Redeneringen verbeteren

Beoordeel de drie uitspraken. Voor a en b gebruik je de oefenreeks 1, 2, 3, 4, 5, 6, 7, 8. Voor c gebruik je 1, 2, 3, 4, 5, 6, 7, 8, 9.

a «Bij acht waarden zijn Q1 en Q3 de tweede en zevende waarde, dus de interkwartielafstand is 7 − 2 = 5.» Verbeter de berekening.

b «Als ik bij elke waarde 5 optel, wordt de standaardafwijking ook 5 groter.» Verbeter de redenering.

c «In de box liggen altijd precies 50% van de waarden. Bij negen waarden zijn dat 4,5 waarden.» Verbeter de redenering met deze reeks.

20 Nieuwe reeks tijden

Samengestelde oefengegevens: tien tijden in minuten zijn 22, 15, 18, 12, 20, 16, 25, 14, 19 en 17.

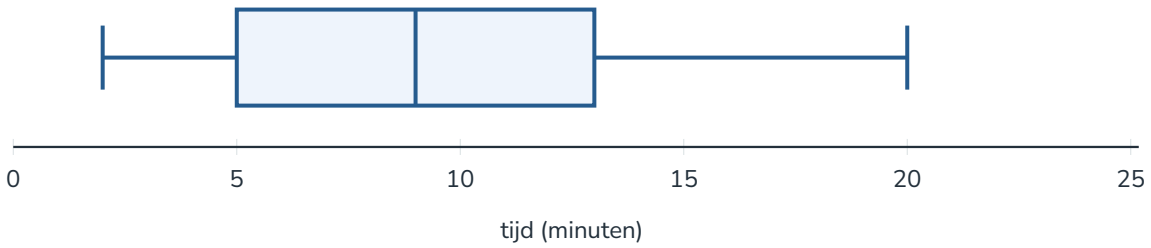
a Bepaal de mediaan, Q1, Q3, spreidingsbreedte en interkwartielafstand.

b Voeg één tijd van 100 minuten toe. Bepaal de nieuwe interkwartielafstand. Verklaar de verandering.

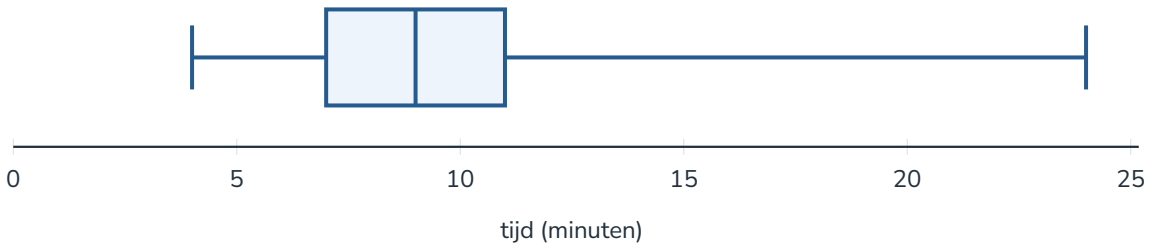
21 Twee boxplots beoordelen

De boxplots horen bij samengestelde oefentijden. Groep B telt twintig verschillende tijden. De handconventie is gebruikt.

Groep A



Groep B



a Welke groep heeft de kleinste spreiding van de middelste helft? Onderbouw met de interkwartielafstanden.

b Hoeveel tijden van groep B liggen tussen 7 en 11 minuten?

c Kun je de standaardafwijking van groep B exact uit de boxplot berekenen? Leg uit.

22 Spreiding na een omzetting

Een oefenreeks scores is 3, 5, 7, 7, 9, 11 punten.

a Bepaal de standaardafwijking σx op twee decimalen.

b Elke score wordt vervangen door $y = 3x - 2$. Bepaal de nieuwe mediaan, interkwartielafstand en standaardafwijking. Rond de standaardafwijking af op twee decimalen.

c Een onderzoeker wil het verschil tussen de twee uiterste scores beschrijven. Welke spreidingsmaat is daarvoor geschikt? Leg uit waarom en bepaal deze maat voor de nieuwe scores.

Herhaling

H1 Bibliotheekbezoek

Een school onderzoekt hoe vaak haar 240 leerlingen uit 4 havo in de afgelopen maand de bibliotheek bezochten. Acht leerlingen reageren op een open oproep. Hun aantallen bezoeken zijn 2, 0, 1, 5, 2, 3, 1 en 2. De onderzoeksgegevens zijn voor deze oefening samengesteld.

a Wat is de populatie? Welk gegevenstype heeft het aantal bezoeken? Leg uit waarom de acht reacties geen aselecte steekproef vormen.

b Maak een frequentietabel met absolute, relatieve en cumulatieve absolute frequenties. Hoeveel procent bezocht de bibliotheek ten hoogste twee keer?

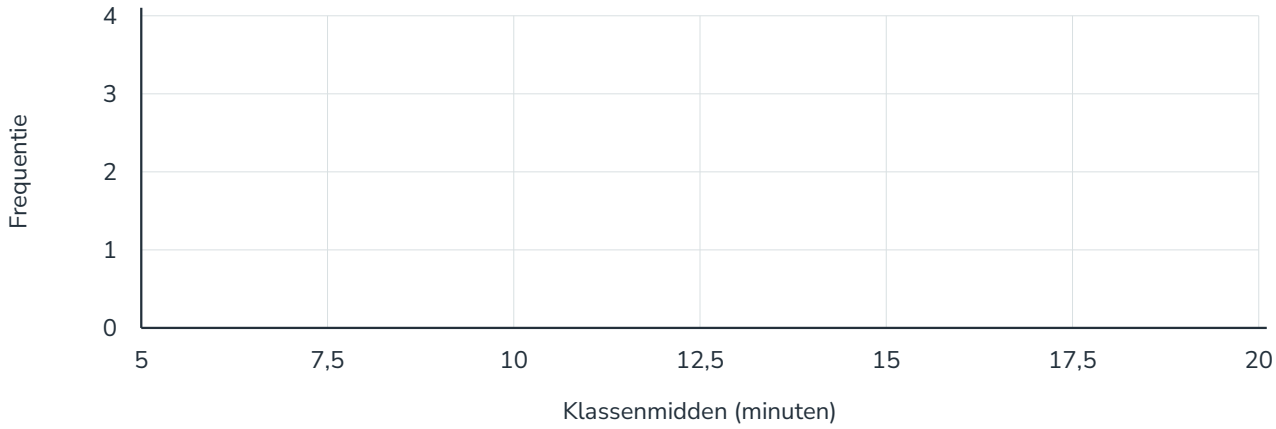
c Bepaal gemiddelde, mediaan en modus. Kies een van die maten voor het aantal bezoeken dat het vaakst voorkomt en verklaar je keuze.

H2 Bezorgtijden

Acht bezorgtijden zijn 6, 8, 8, 10, 10, 12, 14 en 16 minuten. Deze onderzoeksgegevens zijn voor de oefening samengesteld.

a Maak de frequentietabel bij $5 \leq t < 10$, $10 \leq t < 15$ en $15 \leq t \leq 20$. Waarom kan het exacte aantal ritten van acht minuten niet meer uit alleen deze klassentabel worden afgelezen?

Bezorgtijden per klasse



b Teken een frequentiepolygoon bij deze klassentabel.

c Bereken het exacte gemiddelde uit de oorspronkelijke tijden. Bereken ook het gemiddelde als elke tijd door het midden van zijn klasse wordt vervangen. Verklaar het verschil.

H3 Twee afhaallocaties

Bij afhaallocatie A zijn vijf klanten onderzocht en bij B vijftien. De onderzoeksgegevens zijn voor deze oefening samengesteld.

Locatie	Aantal klanten	Gemiddeld bedrag	Tevreden
A	5	€ 12	4
B	15	€ 20	9

a Bereken het gemiddelde aankoopbedrag van alle twintig klanten. Waarom is het gewone gemiddelde van de twee locatiegemiddelden hier ongeschikt?

Tevreden klanten



b Teken een staafdiagram waarmee de percentages tevreden klanten in A en B vergeleken kunnen worden.

c Een bericht noemt B beter, omdat daar negen klanten tevreden waren en in A maar vier. Beoordeel deze onderbouwing.

H4 Waterpeilen

Op acht dagen zijn waterpeilen van 1, 2, 3, 4, 5, 6, 7 en 8 cm geregistreerd. De onderzoeksgegevens zijn voor deze oefening samengesteld. Bij de standaardafwijking mag de grafische rekenmachine worden gebruikt.

Waterpeilen



a Bepaal de vijf waarden voor een boxplot en teken die boxplot. Bepaal ook σx, afgerond op twee decimalen.

b Alle waterpeilen worden omgerekend met z = 2x + 5. Geef het nieuwe gemiddelde, de nieuwe mediaan, de spreidingsbreedte, de interkwartielafstand en de standaardafwijking. Rond de standaardafwijking af op twee decimalen.

c Ga terug naar de oorspronkelijke peilen. De waarde 8 wordt vervangen door 24 cm. Bepaal de nieuwe interkwartielafstand, spreidingsbreedte en standaardafwijking. Welke maat reageert hier het minst op de uitschieter?

Herstel

R1 Een bruikbare steekproef

Een school met 300 leerlingen in 4 en 5 havo wil weten hoeveel tijd deze leerlingen op een schooldag gemiddeld aan huiswerk besteden. Plan A gebruikt twintig vrijwilligers uit een huiswerkclub. Plan B kiest met loting dertig namen uit de volledige leerlingenlijst. Deze situatie is voor de oefening samengesteld.

a Baken de populatie af en noem één variabele die rechtstreeks nodig is om de onderzoeksvraag te beantwoorden.

b Welk plan heeft een aselecte selectie? Geef bij het andere plan een mogelijke oorzaak van vertekening. Garandeert de loting dat de uiteindelijke reacties representatief zijn?

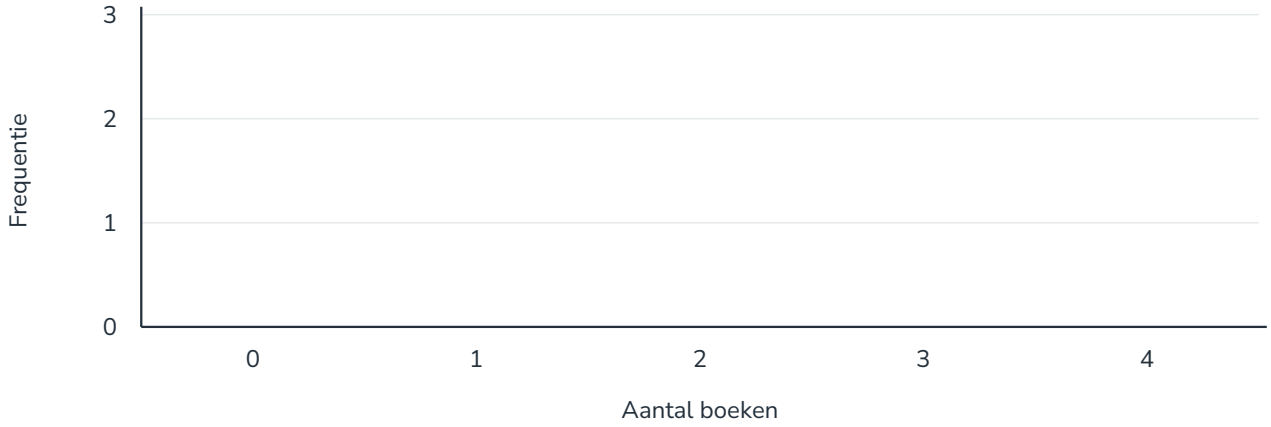
c De school registreert ook leerjaar, tevredenheid (zeer laag tot zeer hoog), aantal gemaakte opdrachten en huiswerktijd afgerond op hele minuten. Deel elk gegeven in bij een gegevenstype en licht je keuze toe.

R2 Van gegevens naar frequenties

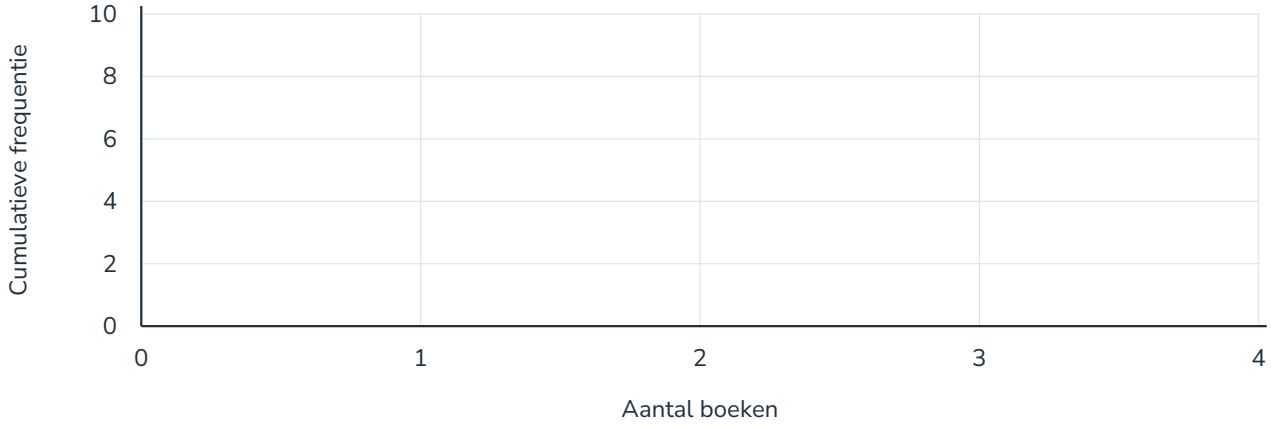
Tien bezoekers leenden afgelopen week 1, 3, 0, 2, 1, 4, 0, 3, 2 en 1 boeken. Deze onderzoeksgegevens zijn voor de oefening samengesteld.

a Maak een tabel met absolute, relatieve en cumulatieve absolute frequenties. Bepaal het percentage bezoekers dat hoogstens twee boeken leende.

Uitgeleende boeken



Cumulatieve uitleenfrequentie



b Teken een staafdiagram van de absolute frequenties en een cumulatief absoluut frequentiepolygoon.

c De gegevens worden samengevat in twee klassen: $0 \leq b < 2$ en $2 \leq b \leq 4$. Geef de klassefrequenties. Kan daarna het exacte aantal bezoekers met één boek uit alleen de twee klassefrequenties worden teruggevonden?

R3 Een lange wachttijd

Zes wachttijden zijn 4, 5, 5, 6, 6 en 16 minuten. Deze onderzoeksgegevens zijn voor de oefening samengesteld. Voor σx is de grafische rekenmachine toegestaan.

a Bepaal gemiddelde, mediaan, modus, kwartielen, interkwartielafstand en spreidingsbreedte.

b Bepaal σx op twee decimalen. Kies een centrummaat en een spreidingsmaat om de gebruikelijke wachttijden te beschrijven en motiveer je keuze.

c De registratie 16 blijkt onjuist en wordt 7 minuten. Welke van de berekende centrum- en spreidingsmaten veranderen? Geef de nieuwe waarden van de veranderde maten; rond σx af op twee decimalen.

Controle

C1 Kruidenpotten op balkons

Een wooncomplex heeft 600 bewoners. Het bestuur wil weten hoeveel kruidenpotten alle bewoners op hun balkon hebben. Acht vrijwilligers uit de tuinclub reageren met de aantallen 1, 3, 2, 1, 4, 2, 3 en 2. Deze onderzoeksgegevens zijn voor de oefening samengesteld.

a Baken de populatie af, deel de meetvariabele in bij een gegevenstype en beoordeel de selectie van de deelnemers.

b Maak een frequentietabel met absolute, relatieve en cumulatieve absolute frequenties. Welk percentage van de acht reacties bevat hoogstens twee potten?

c Kies een diagram om de absolute frequenties van de aantallen potten te tonen. Motiveer je keuze en teken het diagram.

d De aantallen worden in twee klassen gezet: $1 \leq p < 3$ en $3 \leq p \leq 5$. Geef de frequenties en leg uit waarom het exacte gemiddelde niet uit alleen die twee frequenties te berekenen is.

C2 Pakketgewichten

Acht pakketgewichten zijn 2, 3, 4, 4, 6, 7, 8 en 10 kg. Deze onderzoeksgegevens zijn voor de oefening samengesteld. Voor de standaardafwijking is de grafische rekenmachine toegestaan.

a Bepaal gemiddelde, mediaan en modus.

Pakketgewichten



b Bepaal de kwartielen, interkwartielafstand en spreidingsbreedte. Teken een boxplot.

c Bepaal σx op twee decimalen. Welke spreidingsmaat uit dit hoofdstuk gebruikt alle acht waarden rechtstreeks?

d Een nieuwe grootheid wordt voor elk oorspronkelijk pakket berekend met z = 1,5x + 2. Geef het gemiddelde, de mediaan, de interkwartielafstand en de standaardafwijking van z. Rond de standaardafwijking af op twee decimalen.

Leerdoelenoverzicht

2.1 Onderzoeksvragen, steekproeven en soorten gegevens

Ik kan bij een onderzoeksvraag de populatie afbakenen en de benodigde meetvariabelen kiezen.

Ik kan beoordelen hoe een steekproef is geselecteerd en uitleggen waardoor een uitkomst vertekend kan zijn.

Ik kan een variabele indelen als nominaal, ordinaal, discreet kwantitatief of continu kwantitatief en mijn keuze uitleggen.

Ik kan de stappen van een onderzoek herkennen en een conclusie formuleren die past bij de verzamelde gegevens.

2.2 Frequentietabellen en klassenindeling

Ik kan ruwe gegevens ordenen in een frequentietabel met correcte totalen.

Ik kan absolute, relatieve en cumulatieve frequenties berekenen en interpreteren.

Ik kan een klassenindeling maken met consistente grenzen en een passende klassebreedte.

Ik kan uitleggen welke informatie verloren gaat bij het groeperen van gegevens.

2.3 Diagrammen kiezen en omzetten

Ik kan een diagram kiezen dat past bij het gegevenstype en de onderzoeksvraag.

Ik kan een frequentietabel omzetten in een staaf-, cirkel- of lijndiagram en een frequentiepolygoon.

Ik kan cumulatieve voorstellingen, steelbladdiagrammen en dotplots uit brondata opstellen en omzetten, en bij voldoende gekoppelde brondata een kruistabel of puntenwolk maken en lezen.

Ik kan een datapresentatie beoordelen op schaal, weggelaten informatie en vergelijkbaarheid.

2.4 Centrummaten berekenen en kiezen

Ik kan gemiddelde, mediaan en modus bepalen uit losse gegevens of een frequentietabel.

Ik kan een gewogen gemiddelde berekenen en uitleggen wanneer een gewoon gemiddelde ongeschikt is.

Ik kan een centrummaat kiezen bij het gegevenstype en de onderzoeksvraag.

Ik kan de invloed van een uitschieter of een bewerking van alle waarnemingen op centrummaten voorspellen.

2.5 Spreiding en boxplots

Ik kan spreidingsbreedte, kwartielen en interkwartielafstand bepalen uit geordende data.

Ik kan een boxplot maken en waarden of informatie over aantallen uit een boxplot afleiden.

Ik kan standaardafwijking met geschikte ICT bepalen en de keuze tussen standaardafwijking, interkwartielafstand en spreidingsbreedte beargumenteren.

Ik kan uitleggen wat een uitschieter of een bewerking van alle waarden doet met spreidingsmaten.

Eindevaluatie

Kijk terug op je antwoorden en berekeningen. Noteer welke vaardigheden je zelfstandig beheerst en waar je nog oefening nodig hebt.

Vergelijk na het nakijken jouw redenering met het modelantwoord. Een andere juiste aanpak kan ook tot een goed antwoord leiden.

Eindevaluatie

Kijk na met de uitwerkingen. Kies bij een fout de bijbehorende opgaven onder Oefen opnieuw. Maak daarna de opgaven onder Controleer zonder hulp.

Toetsvraag	Leerdoel	Oefen opnieuw	Controleer
Test jezelf T1a–c	Populatie, steekproef en gegevenstype	Herstel R1a–c	Controle C1a
Test jezelf T3c	Onderzoekscyclus en conclusie zonder causaliteitsclaim	2.1 opgave 9a–b; 2.1 opgave 10a–b	2.1 opgave 14b–c
Test jezelf T1d; Test jezelf T2a–c; Test jezelf T5b	Frequenties, klassen en informatieverlies	Herstel R2a,c; 2.2 opgave 11a–b	Controle C1b,d
Test jezelf T1e; Test jezelf T2e; Test jezelf T3a–b; Test jezelf T5a–b	Diagramkeuze en diagrammen uit brondata	Herstel R2b; 2.3 opgave 13a–c; 2.3 opgave 16a–b; 2.3 opgave 17a–b	Controle C1c; 2.3 opgave 23a–e
2.3 opgave 22a–c (aanvullend op Test jezelf)	Cirkel-, lijn- en histogramconstructie	2.3 opgave 6a–d; 2.3 opgave 7a–b	2.3 opgave 9a–b; 2.3 opgave 8a–b
Test jezelf T3d	Misleidende schaal en vergelijkbaarheid	2.3 opgave 19a–b; 2.3 opgave 20a–c	2.3 opgave 24a–c
Test jezelf T2d; Test jezelf T1e; Test jezelf T4c–d	Centrummaten berekenen, kiezen en uitschieter beoordelen	Herstel R3a–c; 2.4 opgave 10a–b	Controle C2a; 2.4 opgave 18a,c
Test jezelf T3e	Gewogen gemiddelde	2.4 opgave 6a–b; 2.4 opgave 9a–b	2.4 opgave 17b
Test jezelf T4a–b; Test jezelf T5c–d	Kwartielen, boxplot, σx en passende spreidingsmaat	Herstel R3a–b; 2.5 opgave 6a–b; 2.5 opgave 8a–b	Controle C2b–c; 2.5 opgave 21a–c
Test jezelf T4e; Test jezelf T5e	Omzettingen en uitschieters bij centrum en spreiding	Herstel R3c; 2.5 opgave 17a–b	Controle C2d; 2.5 opgave 22a–c

Voorbeeld 2.1A

Een wens van sportleden

Een sportcentrum heeft 420 leden. Het wil weten welk percentage van alle leden op zondagochtend wil trainen. Op donderdagavond vullen 32 aanwezige leden vrijwillig een formulier in; 20 antwoorden ja. De onderzoeksgegevens zijn voor dit voorbeeld samengesteld.

a Benoem de populatie en kies een variabele waarmee de vraag kan worden beantwoord.

1. Wat zoek ik?

De volledige doelgroep en een bruikbaar meetgegeven.

2. Welke basis en aanpak kies ik, en waarom?

Ik gebruik de doelgroep uit de vraag. Een directe ja/nee-vraag meet de wens.

3. Berekening

Er is geen rekenbewerking nodig: 420 leden vormen de doelgroep.

4. Antwoord in context

Alle 420 leden; per lid de wens voor een zondagochtendtraining.

5. Controle

Ook een lid dat donderdag niet traint, behoort tot de populatie.

Voorbeeld 2.1A

Een wens van sportleden

Een sportcentrum heeft 420 leden. Het wil weten welk percentage van alle leden op zondagochtend wil trainen. Op donderdagavond vullen 32 aanwezige leden vrijwillig een formulier in; 20 antwoorden ja. De onderzoeksgegevens zijn voor dit voorbeeld samengesteld.

b Bereken het percentage ja onder de respondenten. Beoordeel de selectie en stel een verbetering voor.

1. Wat zoek ik?

Het percentage bij de respondenten en de kwaliteit van de selectie.

2. Welke basis en aanpak kies ik, en waarom?

De 32 respondenten zijn de juiste percentagebasis; alleen zij hebben geantwoord.

3. Berekening

$20 \div 32 \times 100 = 62{,}5\%$.

4. Antwoord in context

62,5% van deze respondenten wil zondagtraining. De selectie kan vertekend zijn. Verbeter de opzet: loot uit de volledige ledenlijst en volg non-respons op.

5. Controle

$0{,}625 \times 32 = 20$. Leden zonder donderdagbezoek zijn niet vanzelf bereikt.

Voorbeeld 2.1B

Reizen naar de sporthal

Uit alle 300 leden van een vereniging worden 40 leden aselect gekozen. Iedereen antwoordt. De vereniging registreert vervoermiddel, tevredenheid over de reis (ontevreden, neutraal, tevreden), aantal trainingen per week en reisafstand op één decimaal. Van de 40 leden reizen er 26 met de fiets. De onderzoeksgegevens zijn voor dit voorbeeld samengesteld.

a Deel de vier geregistreerde variabelen in naar soort.

1. Wat zoek ik?

Het soort van elke variabele.

2. Welke basis en aanpak kies ik, en waarom?

Ik vraag per gegeven: categorie of hoeveelheid; daarna rangorde, telling of meting.

3. Berekening

Geen rekenbewerking: de soorten volgen uit de betekenis van de gegevens.

4. Antwoord in context

De vier soorten zijn respectievelijk nominaal, ordinaal, discreet en continu kwantitatief.

5. Controle

Een afstand van 4,3 km is afgerond; de werkelijke afstand kan ook tussenwaarden hebben.

Voorbeeld 2.1B

Reizen naar de sporthal

Uit alle 300 leden van een vereniging worden 40 leden aselect gekozen. Iedereen antwoordt. De vereniging registreert vervoermiddel, tevredenheid over de reis (ontevreden, neutraal, tevreden), aantal trainingen per week en reisafstand op één decimaal. Van de 40 leden reizen er 26 met de fiets. De onderzoeksgegevens zijn voor dit voorbeeld samengesteld.

b Welke twee onderzoeksstappen volgen na data verzamelen? Werk ze uit met het aantal fietsers.

1. Wat zoek ik?

Een analyse en een passende conclusie.

2. Welke basis en aanpak kies ik, en waarom?

Ik gebruik 40 als basis, omdat van die 40 het vervoermiddel bekend is.

3. Berekening

$26 \div 40 \times 100 = 65\%$.

4. Antwoord in context

65% van de 40 deelnemers reist met de fiets.

5. Controle

$0{,}65 \times 40 = 26$. Uit de andere 260 leden zijn geen vervoergegevens verzameld.

Voorbeeld 2.2A

Van ruwe waarden naar een tabel

Oefengegevens: 12 personen noteren het aantal overstappen tijdens een reis: 0, 1, 0, 2, 1, 1, 3, 2, 1, 0, 2, 1.

a Maak een frequentietabel en controleer het totaal.

1. Wat zoek ik?

Het aantal reizen bij elke waarde voor het aantal overstappen.

2. Welke basis en aanpak kies ik, en waarom?

Een geordende frequentietabel past bij de getalswaarden. Ik tel elke registratie precies eenmaal.

3. Berekening

De tellingen zijn 3, 5, 3 en 1.

Aantal overstappen	Absolute frequentie
0	3
1	5
2	3
3	1
Totaal	12

4. Antwoord in context

Drie reizen hebben geen overstap, vijf hebben er één, drie hebben er twee en één heeft er drie.

5. Controle

3 + 5 + 3 + 1 = 12; bij opnieuw tellen zie ik vijf enen.

Voorbeeld 2.2A

Van ruwe waarden naar een tabel

Oefengegevens: 12 personen noteren het aantal overstappen tijdens een reis: 0, 1, 0, 2, 1, 1, 3, 2, 1, 0, 2, 1.

b Voeg relatieve frequenties in procenten en cumulatieve absolute frequenties toe. Welk percentage van deze reizen had hoogstens één overstap? Rond percentages af op één decimaal.

1. Wat zoek ik?

Percentages per waarde, cumulatieve aantallen en het percentage bij hoogstens één.

2. Welke basis en aanpak kies ik, en waarom?

Het totaal is 12 reizen. Omdat overstappen een natuurlijke volgorde hebben, kan ik de frequenties cumuleren.

3. Berekening

De cumulatieve frequentie bij 1 is 3 + 5 = 8; $\frac{8}{12} \times 100\% \approx 66{,}7\%$.

Aantal overstappen	Absolute frequentie	Relatieve frequentie	Cumulatieve absolute frequentie
0	3	25,0%	3
1	5	41,7%	8
2	3	25,0%	11
3	1	8,3%	12

4. Antwoord in context

Ongeveer twee derde van deze reizen had hoogstens één overstap.

5. Controle

De laatste cumulatieve frequentie is 12. De percentages zijn samen 100,0%, en 66,7% ligt tussen 0% en 100%.

Voorbeeld 2.2B

Groeperen en informatie behouden

Oefengegevens: laadtijden t in seconden zijn 4,8; 5,0; 7,3; 9,9; 10,0; 11,6; 14,9; 15,0; 16,4; 18,7; 19,9; 20,0.

a Maak een frequentietabel met startgrens 0 s en klassebreedte 5 s. Kies de laatste bovengrens boven de grootste meting.

1. Wat zoek ik?

Klassen van 5 seconden en de aantallen in die klassen.

2. Welke basis en aanpak kies ik, en waarom?

Ik begin bij 0, verhoog grenzen steeds met 5 en gebruik ondergrens inclusief, bovengrens exclusief. Zo telt een grenswaarde één keer.

3. Berekening

De grootste meting is 20,0; daarom is ook $20 \le t < 25$ nodig. De frequenties zijn 1, 3, 3, 4 en 1.

Laadtijd t (s)	Absolute frequentie
$0 \le t < 5$	1
$5 \le t < 10$	3
$10 \le t < 15$	3
$15 \le t < 20$	4
$20 \le t < 25$	1
Totaal	12

4. Antwoord in context

De klassentabel laat zien hoeveel laadtijden in elk interval liggen.

5. Controle

$1 + 3 + 3 + 4 + 1 = 12$. De waarden 5,0; 10,0; 15,0 en 20,0 staan steeds in de klasse die daar begint.

Voorbeeld 2.2B

Groeperen en informatie behouden

Oefengegevens: laadtijden t in seconden zijn 4,8; 5,0; 7,3; 9,9; 10,0; 11,6; 14,9; 15,0; 16,4; 18,7; 19,9; 20,0.

b Je krijgt alleen de klassentabel. Vergelijk wat je kunt zeggen over het aantal laadtijden korter dan 15 s en het aantal korter dan 12 s.

1. Wat zoek ik?

Welke van de twee aantallen de klassentabel precies bepaalt.

2. Welke basis en aanpak kies ik, en waarom?

Ik kijk of de gevraagde grens samenvalt met een klassegrens. Volledige klassen kan ik zeker meetellen.

3. Berekening

Onder 15: 1 + 3 + 3 = 7. Onder 12: 1 + 3 plus 0 tot 3, dus 4 tot en met 7.

4. Antwoord in context

De tabel bewaart het exacte aantal onder 15, maar verbergt het exacte aantal onder 12.

5. Controle

Alle drie waarden uit $10 \leq t < 15$ kunnen bijvoorbeeld 10,5 zijn of 14,5. Beide mogelijkheden houden de tabel gelijk en geven de twee uiterste aantallen.

Voorbeeld 2.3A

Kiezen vanuit de vraag

De gegevens van deze twee kleine onderzoeken zijn voor dit voorbeeld samengesteld.

Tabel I: workshop	Aantal
Koken	16
Tekenen	12
Muziek	8
Fotografie	4

a Tabel I geeft de gekozen workshop. Je wilt de aantallen gemakkelijk vergelijken. Welk diagram kies je? Leg uit waarom een lijn hier niet past.

1. Wat zoek ik?

Welke workshop vaker gekozen wordt.

2. Welke basis en aanpak kies ik, en waarom?

Ik kies losse staven, omdat ik categorieën op grootte wil vergelijken.

3. Berekening

De aantallen zijn 16, 12, 8 en 4; de grootste staaf hoort bij koken.

4. Antwoord in context

Een staafdiagram laat het verschil tussen de workshops zien.

5. Controle

Er is geen tijdas of numerieke workshopas; verbonden punten zouden een volgorde suggereren.

Voorbeeld 2.3A

Kiezen vanuit de vraag

De gegevens van deze twee kleine onderzoeken zijn voor dit voorbeeld samengesteld.

Tabel II: week	Inschrijvingen
1	8
2	12
3	10
4	16

b Tabel II geeft per week het aantal inschrijvingen. Je wilt de ontwikkeling volgen. Welk diagram kies je, en wat worden de assen?

1. Wat zoek ik?

De ontwikkeling gedurende vier weken.

2. Welke basis en aanpak kies ik, en waarom?

Een tijdreeks heeft een vaste volgorde; daarom kan ik opeenvolgende meetpunten verbinden.

3. Berekening

De punten zijn (1, 8), (2, 12), (3, 10) en (4, 16).

4. Antwoord in context

Het lijndiagram laat eerst een stijging, dan een daling en weer een stijging zien.

5. Controle

De verticale as heeft de eenheid aantal inschrijvingen; de horizontale as geeft weken.

Voorbeeld 2.3B

Aantallen en aandelen tekenen

Twintig bezoekers kiezen één kaartontwerp. Deze voorbeeldgegevens zijn samengesteld.

Ontwerp	Frequentie
Rond	5
Vierkant	9
Driehoek	6

a Teken een staafdiagram van de aantallen.

1. Wat zoek ik?

De aantallen per ontwerp in een staafdiagram.

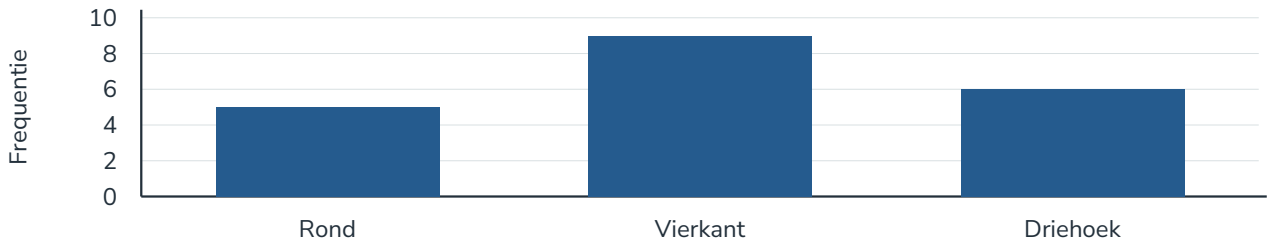
2. Welke basis en aanpak kies ik, en waarom?

Voor nominale categorieën gebruik ik losse staven met dezelfde breedte.

3. Berekening

De hoogten zijn de gegeven frequenties: 5, 9 en 6.

Kaartontwerpen



4. Antwoord in context

De staaf van vierkant is het hoogst.

5. Controle

5 + 9 + 6 = 20; de verticale as heet frequentie en begint bij 0.

Voorbeeld 2.3B

Aantallen en aandelen tekenen

Twintig bezoekers kiezen één kaartontwerp. Deze voorbeeldgegevens zijn samengesteld.

Ontwerp	Frequentie
Rond	5
Vierkant	9
Driehoek	6

b Teken een cirkeldiagram. Bereken eerst voor ieder ontwerp de sectorhoek.

1. Wat zoek ik?

Aandelen van de 20 keuzes.

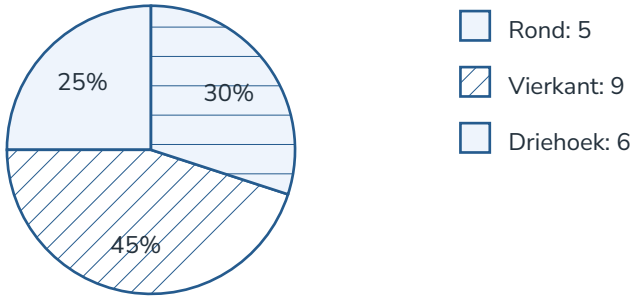
2. Welke basis en aanpak kies ik, en waarom?

Eén cirkel is het geheel van 20 keuzes.

3. Berekening

De hoeken zijn 90°, 162° en 108°.

Aandelen kaartontwerpen



4. Antwoord in context

Vierkant beslaat 45% van de cirkel.

5. Controle

De hoeken tellen op tot 360°; de percentages tot 100%.

Voorbeeld 2.3C

Brongegevens op drie manieren gebruiken

De voorbeeldgegevens zijn samengesteld. a en b gebruiken aparte datasets; c en d dezelfde scoreparen.

Tabel I: wachttijd t (min)	Frequentie
$0 \le t < 10$	2
$10 \le t < 20$	3
$20 \le t \le 30$	3

a Maak bij tabel I de cumulatieve tabel en procentcurve. Welk percentage wacht minder dan 20 minuten?

1. Wat zoek ik?

Het aandeel onder 20 minuten.

2. Welke basis en aanpak kies ik, en waarom?

Tel op in klassevolgorde; de basis is 8.

Voorbeeld 2.3C (vervolg)

Brongegevens op drie manieren gebruiken

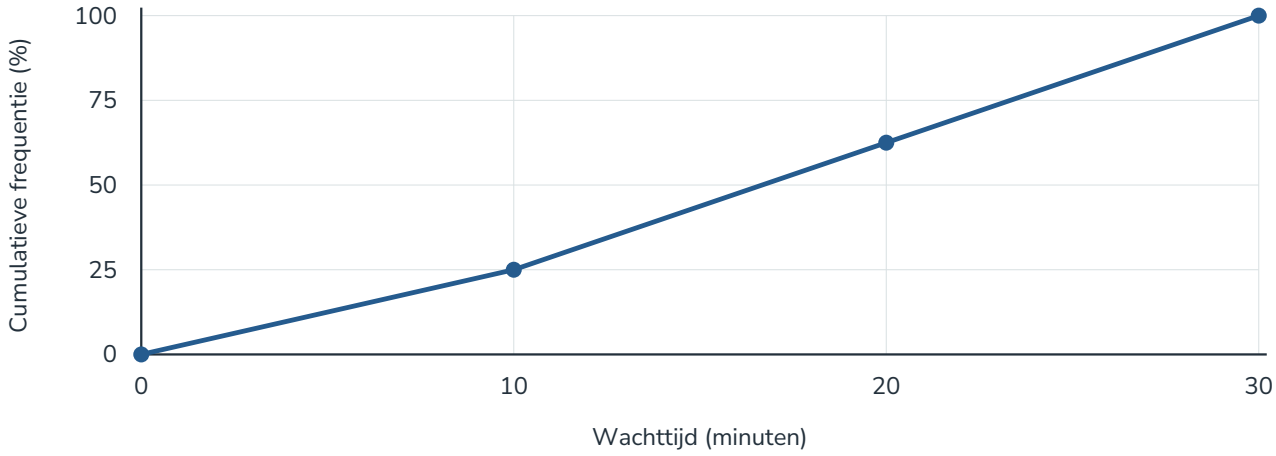
a Maak bij tabel I de cumulatieve tabel en procentcurve. Welk percentage wacht minder dan 20 minuten?

3. Berekening

$\frac{2}{8}$ = 25%; $\frac{5}{8}$ = 62,5%; $\frac{8}{8}$ = 100%.

Bovengrens (min)	Cumulatief aantal	Cumulatief %
10	2	25
20	5	62,5
30	8	100

Cumulatieve wachttijd



4. Antwoord in context

Bij 20 minuten lees ik 62,5%.

5. Controle

De curve begint op 0 en eindigt op 100%.

Voorbeeld 2.3C

Brongegevens op drie manieren gebruiken

De voorbeeldgegevens zijn samengesteld. a en b gebruiken aparte datasets; c en d dezelfde scoreparen.

b De zes andere wachttijden zijn 11, 12, 12, 14, 15 en 15 minuten. Maak een steelbladdiagram en een dotplot.

1. Wat zoek ik?

De afzonderlijke tijden, inclusief herhalingen.

2. Welke basis en aanpak kies ik, en waarom?

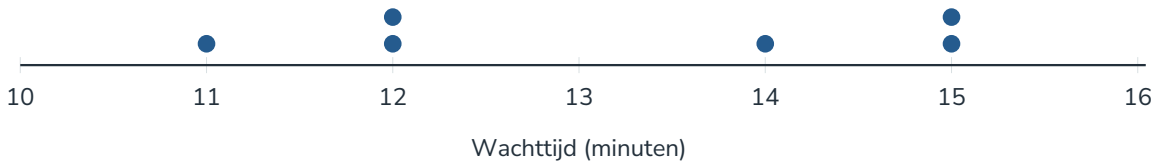
Een steelbladdiagram bewaart de getallen; een dotplot laat de herhalingen op één as zien.

3. Berekening

De stam is steeds 1; de bladeren zijn 1, 2, 2, 4, 5, 5.

Stam	Bladeren
1	1 2 2 4 5 5

Zes wachttijden



4. Antwoord in context

Sleutel: 1 | 2 betekent 12 minuten. Bij 12 en 15 minuten staan elk twee stippen.

5. Controle

Ik tel zes bladeren en zes stippen; ze geven dezelfde zes tijden.

Voorbeeld 2.3C

Brongegevens op drie manieren gebruiken

De voorbeeldgegevens zijn samengesteld. a en b gebruiken aparte datasets; c en d dezelfde scoreparen.

Tabel II: leerling	Score x	Score y
1	1	3
2	2	5
3	3	5
4	4	7

c Tabel II geeft vier gekoppelde scoreparen. Maak een kruistabel met rijen $x < 3$ en $x \geq 3$, kolommen $y < 6$ en $y \geq 6$.

1. Wat zoek ik?

De aantallen per combinatie van scoregroepen.

2. Welke basis en aanpak kies ik, en waarom?

Houd de scores van iedere leerling gekoppeld.

3. Berekening

De kruisfrequenties zijn 2, 0, 1 en 1.

Scoregroep x	$y < 6$	$y \geq 6$	Totaal
$x < 3$	2	0	2
$x \geq 3$	1	1	2
Totaal	3	1	4

4. Antwoord in context

Drie leerlingen hebben $y < 6$; één heeft $y \geq 6$.

5. Controle

De rij- en kolomtotalen tellen beide op tot vier.

Voorbeeld 2.3C

Brongegevens op drie manieren gebruiken

De voorbeeldgegevens zijn samengesteld. a en b gebruiken aparte datasets; c en d dezelfde scoreparen.

Tabel II: leerling	Score x	Score y
1	1	3
2	2	5
3	3	5
4	4	7

d Teken de puntenwolk van de scoreparen in tabel II. Beschrijf het patroon.

1. Wat zoek ik?

Het patroon van de vier scoreparen.

2. Welke basis en aanpak kies ik, en waarom?

Twee kwantitatieve scores per leerling passen in een puntenwolk.

Voorbeeld 2.3C (vervolg)

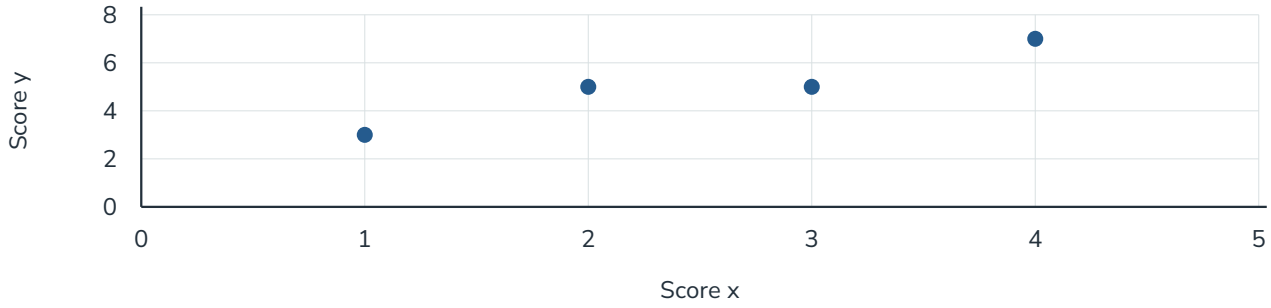
Brongegevens op drie manieren gebruiken

d Teken de puntenwolk van de scoreparen in tabel II. Beschrijf het patroon.

3. Berekening

De punten zijn (1, 3), (2, 5), (3, 5), (4, 7).

Gekoppelde scores



4. Antwoord in context

Grotere x gaat hier meestal samen met grotere y.

5. Controle

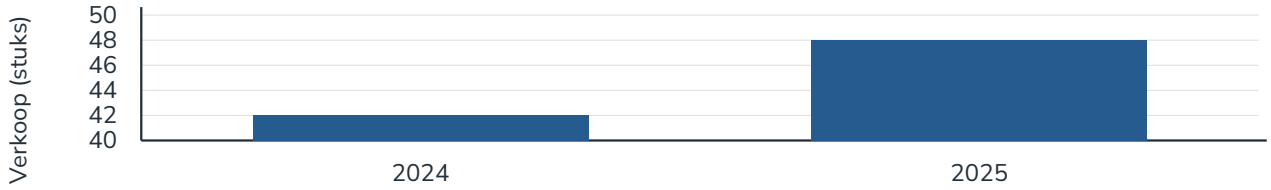
Er zijn vier bronparen en vier punten; dit bewijst geen oorzaak.

Voorbeeld 2.3D

Een eerlijke vergelijking

Deze verkoop- en groepsgegevens zijn voor het voorbeeld samengesteld.

Verkoop volgens de advertentie



a De verkoop stijgt van 42 naar 48 stuks. In de figuur lijkt de tweede staaf viermaal zo hoog. Beoordeel die indruk en teken een verbeterd staafdiagram.

1. Wat zoek ik?

Of de hoogteverhouding de verhouding van de verkoop weergeeft.

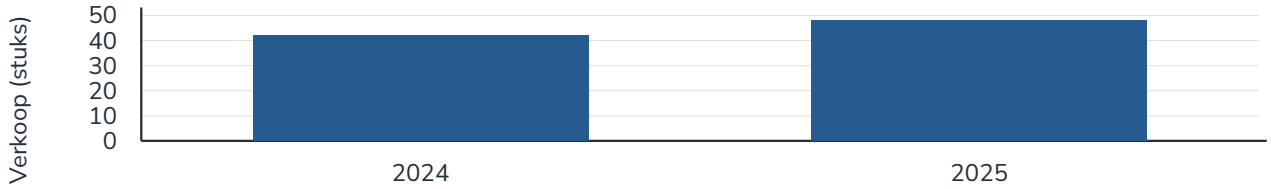
2. Welke basis en aanpak kies ik, en waarom?

Ik lees eerst de as en vergelijk daarna de echte aantallen.

3. Berekening

$(48 - 42)/42 \times 100 \approx 14{,}3\%$; de zichtbare hoogteverhouding is $\frac{8}{2} = 4$.

Verkoop met volledige staafhoogten



4. Antwoord in context

Het diagram met as vanaf 0 toont een bescheiden stijging.

5. Controle

$\frac{48}{42} \approx 1{,}14$, niet 4; beide verbeterde staven gebruiken dezelfde schaal.

Voorbeeld 2.3D

Een eerlijke vergelijking

Deze verkoop- en groepsgegevens zijn voor het voorbeeld samengesteld.

b Groep A heeft 12 geslaagden van 20 deelnemers; groep B 18 van 40. Welke groep heeft het grootste aandeel geslaagden? Maak de vergelijking zichtbaar.

1. Wat zoek ik?

Welke groep naar verhouding meer geslaagden heeft.

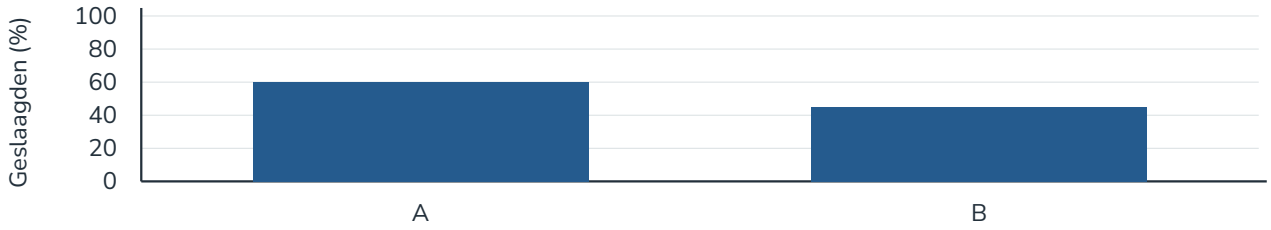
2. Welke basis en aanpak kies ik, en waarom?

De groepen zijn niet even groot; daarom vergelijk ik percentages.

3. Berekening

Voor A deel ik door 20, voor B door 40. Dat geeft 60% en 45%.

Aandeel geslaagden



4. Antwoord in context

Het percentage van A is 15 procentpunten hoger.

5. Controle

Beide percentages vallen tussen 0% en 100%; de noemers zijn de eigen groepsgroottes.

Voorbeeld 2.4A

Drie maten uit losse tijden

Zes oefentijden zijn 6, 4, 5, 4, 9 en 8 minuten. De gegevens zijn voor dit voorbeeld samengesteld.

a Bereken het gemiddelde en bepaal de modus.

1. Wat zoek ik?

Het gemiddelde en de vaakst voorkomende tijd.

2. Welke basis en aanpak kies ik, en waarom?

Voor het gemiddelde gebruik ik alle zes tijden; voor de modus tel ik gelijke tijden.

3. Berekening

(6 + 4 + 5 + 4 + 9 + 8) : 6 = 6. De frequentie van 4 is 2.

4. Antwoord in context

De gemiddelde oefentijd is 6 minuten; de modus is 4 minuten.

5. Controle

6 × 6 = 36 herstelt de totale tijd. Alleen 4 komt vaker dan eenmaal voor.

Voorbeeld 2.4A

Drie maten uit losse tijden

Zes oefentijden zijn 6, 4, 5, 4, 9 en 8 minuten. De gegevens zijn voor dit voorbeeld samengesteld.

b Bepaal de mediaan.

1. Wat zoek ik?

De middelste oefentijd van een even aantal metingen.

2. Welke basis en aanpak kies ik, en waarom?

Ik orden eerst, want de mediaan hangt van de centrale plaatsen af.

3. Berekening

4, 4, 5, 6, 8, 9; (5 + 6) : 2 = 5,5.

4. Antwoord in context

De mediaan van de oefentijden is 5,5 minuten.

5. Controle

Er liggen drie tijden onder en drie boven 5,5; 5,5 ligt tussen de twee centrale tijden.

Voorbeeld 2.4B

Groepen combineren en eenheden omrekenen

Bij meetpunt A zijn drie temperaturen gemeten en bij B twee. De gegevens zijn voor dit voorbeeld samengesteld. Om Celsius om te rekenen naar Fahrenheit geldt F = 1,8C + 32.

Meetpunt	Metingen in °C	Aantal	Gemiddelde in °C
A	12; 14; 16	3	14
B	18; 20	2	19

a Bereken het gemiddelde van alle vijf temperaturen.

1. Wat zoek ik?

Het gemiddelde van alle waarnemingen samen.

2. Welke basis en aanpak kies ik, en waarom?

Ik gebruik 3 en 2 als gewichten, omdat de meetpunten niet evenveel metingen hebben.

3. Berekening

$(3 \times 14 + 2 \times 19) : (3 + 2) = 80 : 5 = 16$.

4. Antwoord in context

De gemiddelde temperatuur is 16 °C.

5. Controle

De losse metingen hebben som $12 + 14 + 16 + 18 + 20 = 80$; $80 : 5 = 16$.

Voorbeeld 2.4B

Groepen combineren en eenheden omrekenen

Bij meetpunt A zijn drie temperaturen gemeten en bij B twee. De gegevens zijn voor dit voorbeeld samengesteld. Om Celsius om te rekenen naar Fahrenheit geldt F = 1,8C + 32.

Meetpunt	Metingen in °C	Aantal	Gemiddelde in °C
A	12; 14; 16	3	14
B	18; 20	2	19

b Bepaal de mediaan in graden Celsius. Voorspel vervolgens gemiddelde en mediaan in graden Fahrenheit.

1. Wat zoek ik?

De mediaan in Celsius en twee centrummaten na omzetting.

2. Welke basis en aanpak kies ik, en waarom?

De omzetting geldt voor elke meting en heeft positieve factor 1,8; ik kan de centrummaten dus ook omzetten.

3. Berekening

Mediaan 16 °C; 1,8 × 16 + 32 = 60,8 °F.

4. Antwoord in context

Gemiddelde en mediaan zijn na omzetting beide 60,8 °F.

5. Controle

De omgerekende waarden zijn 53,6; 57,2; 60,8; 64,4; 68 °F. Hun gemiddelde is 304 : 5 = 60,8; hun middelste waarde is 60,8.

Voorbeeld 2.5A

Kwartielen en een boxplot

Voor deze oefening zijn negen zoektijden samengesteld: 4, 5, 6, 7, 8, 10, 12, 13 en 15 minuten.

a Bepaal de kwartielen, spreidingsbreedte en interkwartielafstand.

1. Wat zoek ik?

De twee kwartielen en twee maten voor spreiding.

2. Welke basis en aanpak kies ik, en waarom?

De reeks is geordend en heeft negen waarden; ik laat de middelste 8 buiten beide helften.

3. Berekening

Q1 = 5,5 en Q3 = 12,5; 15 − 4 = 11 en 12,5 − 5,5 = 7.

4. Antwoord in context

De totale spreiding is 11 min; die van de middelste helft is 7 min.

5. Controle

$4 \le 5{,}5 \le 8 \le 12{,}5 \le 15$ en $7 \le 11$; de eenheid blijft minuten.

Voorbeeld 2.5A

Kwartielen en een boxplot

Voor deze oefening zijn negen zoektijden samengesteld: 4, 5, 6, 7, 8, 10, 12, 13 en 15 minuten.

b Maak de boxplot.

1. Wat zoek ik?

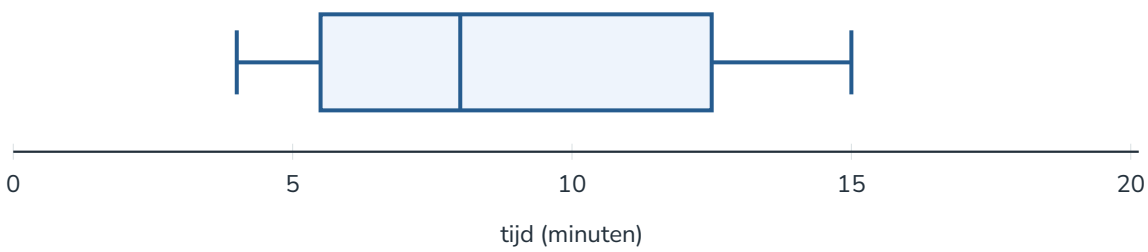
Een boxplot van de negen zoektijden.

2. Welke basis en aanpak kies ik, en waarom?

Ik gebruik minimum, Q1, mediaan, Q3 en maximum op één schaal.

3. Berekening

De box loopt van 5,5 tot 12,5 en de mediaanlijn staat bij 8; de uiteinden staan bij 4 en 15.



4. Antwoord in context

De figuur toont zoektijden in minuten.

5. Controle

De boxlengte is 7 en de totale lengte 11; dat past bij de berekende spreidingsmaten.

Voorbeeld 2.5B

σx lezen en gebruiken

Een oefenreeks scores is 6, 8, 10, 12, 14 punten.

a Bepaal het gemiddelde en de standaardafwijking σx op twee decimalen.

1. Wat zoek ik?

Het gemiddelde en de spreiding van de vijf scores rond dat gemiddelde.

2. Welke basis en aanpak kies ik, en waarom?

Ik gebruik éénvariabele statistiek; de beschreven dataset vraagt σx met deler n.

3. Berekening

De GR geeft n = 5, gemiddelde = 10 en σx = 2,828427… .

4. Antwoord in context

Het gemiddelde is 10 punten en de standaardafwijking 2,83 punten.

5. Controle

De som van de gekwadrateerde afwijkingen is 40; $\sqrt{(\frac{40}{5})}$ geeft dezelfde uitkomst.

Voorbeeld 2.5B

σx lezen en gebruiken

Een oefenreeks scores is 6, 8, 10, 12, 14 punten.

b Liggen alle scores op hoogstens één standaardafwijking van het gemiddelde?

1. Wat zoek ik?

Of elke score hoogstens σx van het gemiddelde af ligt.

2. Welke basis en aanpak kies ik, en waarom?

Ik vergelijk de afzonderlijke afstanden met σx; σx is geen maximale afstand.

3. Berekening

De uiterste scores liggen 4 punten van 10; 4 is groter dan 2,828427… .

4. Antwoord in context

Twee scores liggen verder weg; drie van de vijf liggen binnen deze afstand.

5. Controle

De afstanden zijn 4, 2, 0, 2 en 4; de conclusie past bij alle vijf waarden.

Voorbeeld 2.5B

σx lezen en gebruiken

Een oefenreeks scores is 6, 8, 10, 12, 14 punten.

c Elke score wordt vervangen door $y = 2x + 5$. Bepaal de nieuwe mediaan, interkwartielafstand en standaardafwijking op twee decimalen.

1. Wat zoek ik?

De drie maten na de omzetting van de scores.

2. Welke basis en aanpak kies ik, en waarom?

De factor 2 verdubbelt afstanden; het optellen van 5 verschuift de scores en hun centrum evenveel.

3. Berekening

Mediaan: $2 \times 10 + 5 = 25$; interkwartielafstand: $2 \times 6 = 12$; $\sigma y = 5{,}656854\ldots$.

4. Antwoord in context

De nieuwe mediaan is 25 punten, interkwartielafstand 12 punten en $\sigma y \approx 5{,}66$ punten.

5. Controle

Nieuwe scores zijn 17, 21, 25, 29, 33. Hun kwartielen zijn 19 en 31; $\sigma y = \sqrt{(\frac{160}{5})} = \sqrt{32}$.