

2. Data en datasets verwerken

Boekje 2 havo wiskunde A, domein E: Statistiek



Verantwoording



© 2015, SLO (nationaal expertisecentrum leerplanontwikkeling), Enschede

Dit lesmateriaal is ontwikkeld in het kader van de nieuwe examenprogramma's zoals voorgesteld door de commissie Toekomst Wiskunde Onderwijs (cTWO) en herzien door SLO.

Mits de bron wordt vermeld, is het toegestaan zonder voorafgaande toestemming van de uitgever deze uitgave geheel of gedeeltelijk te kopiëren en/of verspreiden en om afgeleid materiaal te maken dat op deze uitgave is gebaseerd.

Auteurs: Erik van Barneveld, Wouter Boer, Carel van de Giessen, Peter Kop, Heleen van der Ree, Henk Reuling, Frits Spijkers, Tanja Stroosma, Anneke Verschut

Met medewerking van: Nico Alink, Martine de Klein (eindredactie)

Informatie: SLO
Afdeling: tweede fase
Postbus 2041, 7500 CA Enschede
Telefoon (053) 4840 661
Internet: www.slo.nl
E-mail: tweedefase@slo.nl



Overzicht lesmateriaal in het domein Statistiek

1. Kijken naar data

- § 1.1 Wat is statistiek?
- § 1.2 Data
- § 1.3 Diagrammen
- § 1.4 Interpretaties
- § 1.5 Overzicht

2. Data en datasets verwerken

- § 2.0 Begrippenlijst
- § 2.1 Data presenteren
- § 2.2 Verbanden tussen datarepresentaties
- § 2.3 Frequentieverdelingen typeren
- § 2.4 Twee groepen vergelijken
- § 2.5 Samenhang tussen twee variabelen

3. Data verwerven

- § 3.0 Pas op voor valkuilen
- § 3.1 Onderzoeks- en enquêtevragen
- § 3.2 Steekproeven en fouten
- § 3.3 Standaardafwijking
- § 3.4 Steekproeffout: variatie bij steekproeven
- § 3.5 Normale verdeling
- § 3.6 Toevallige steekproeffouten in getallen
- § 3.7 Terugblik op boekje 3

4. Statistische uitspraken doen

- § 4.1 Voorkennis
- § 4.2 Doel van deze module
- § 4.3 Populatieproportie
- § 4.4 Populatiegemiddelde
- § 4.5 Verschil tussen twee groepen
- § 4.6 Samenhang tussen twee kwantitatieve variabelen
- § 4.7 Gemengde opgaven
- § 4.8 Terugblik
- § 4.9 Lessenserie: Statistiek op een groot gegevensbestand
- § 4.10 Diagnostische computertoets



Inhoud

Overzicht lesmateriaal in het domein Statistiek.....	3
§ 2.0 Begrippenlijst.....	5
§ 2.1 Data presenteren	9
§ 2.1.1 Introductie.....	9
§ 2.1.2 Centrale vraag	9
§ 2.1.3 Representaties.....	10
§ 2.1.4 Oefenen.....	16
§ 2.1.5 Om te onthouden	21
§ 2.1.6 Geïntegreerd oefenen.....	22
§ 2.2 Verbanden tussen datarepresentaties	24
§ 2.2.1 Introductie.....	24
§ 2.2.2 Centrale vraag	24
§ 2.2.3 Verbanden tussen representaties	25
§ 2.2.4 Oefenen.....	29
§ 2.2.5 Om te onthouden	31
§ 2.2.6 Geïntegreerd oefenen.....	32
§ 2.3 Frequentieverdelingen typeren.....	35
§ 2.3.1 Introductie.....	35
§ 2.3.2 Centrale vraag	35
§ 2.3.3 Frequentieverdelingen typeren	36
§ 2.3.4 Oefenen.....	42
§ 2.3.5 Om te onthouden	54
§ 2.3.6 Geïntegreerd oefenen.....	54
§ 2.4 Twee groepen vergelijken	56
§ 2.4.1 Introductie.....	56
§ 2.4.2 Centrale vraag	56
§ 2.4.3 Twee groepen vergelijken.....	57
§ 2.4.4 Oefenen.....	61
§ 2.4.5 Om te onthouden	65
§ 2.4.6 Geïntegreerd oefenen.....	65
§ 2.5 Samenhang tussen twee variabelen	66
§ 2.5.1 Introductie.....	66
§ 2.5.2 Centrale vraag	66
§ 2.5.3 Samenhang	67
§ 2.5.4 Oefenen.....	70
§ 2.5.5 Om te onthouden	73
§ 2.5.6 Geïntegreerd oefenen.....	74



§ 2.0 Begrippenlijst

Hieronder zie je een lijst van begrippen die je tegenkomt in deze paragraaf.

Opgave 0

Neem de begrippen in deze lijst door en noteer welke je al kent en welke nieuw voor je zijn.

Tijdens een statistisch onderzoek onderzoek je de (steekproef)populatie op bepaalde kenmerken.

Dataset	Een verzameling waarden van een of meer statistische variabelen.
Deelgroep	Groep variabelen uit de dataset die een of meer dezelfde kenmerken hebben en apart onderzocht worden (bijv. om ze met elkaar te vergelijken).
Variabele	Ieder te onderzoeken kenmerk leg je vast in een statistische variabele die verschillende waarden kan aannemen.
Kwalitatieve variabele	Variabelen als geslacht, kleur ogen, godsdienst, bloedgroep, naam, e.d. Ze geven alleen een kenmerk van de populatie weer, maar hoeven niet noodzakelijkerwijs in een getal te worden uitgedrukt.
Nominale variabele	Variabelen met namen als waarde. Er zit geen volgorde in de waarden.
Ordinale variabele	Er zit een volgorde in de mogelijke waarden, bijv. de mate van interesse in een onderwerp die kan oplopen van geen, een beetje, gemiddeld, redelijk en veel.
Kwantitatieve variabele	Variabelen als lengte, hoogte van het inkomen, omvang van het gezin, e.d. Deze moeten wel in een getal worden uitgedrukt.
Continue variabele	Variabelen zoals lengte, gewicht e.d. Deze kunnen alle waarden binnen een interval aannemen.
Discrete variabele	Variabelen zoals omvang van gezin, aantal ogen bij een worp met een dobbelsteen e.d. Deze nemen alleen bepaalde waarden aan.

Voor iedere statistische variabele houd je bij welke waarden van dit kenmerk in de (steekproef)populatie voorkomen en hoe vaak.

Frequentie	Aantal keren dat een waarde van een variabele voorkomt in een steekproefpopulatie.
Absolute frequentie	De frequentie leg je vast als het aantal dat je in de populatie geteld hebt.
Relatieve frequentie	De frequentie leg je vast als percentage van de totale steekproefomvang.
Cumulatieve frequentie (somfrequentie)	De som van de frequenties van een waarde en alle kleinere waarden. Je stapelt dan als het ware de frequenties op elkaar: bij elke frequentie tel je die van de voorgaande waarden op (alleen bij kwantitatieve variabelen).
Frequentietabel	Tabel over een statistische variabele. Per waarde staat de bijbehorende frequentie.
Frequentieverdeling	Overzicht van het aantal keren dat de verschillende waarden van een variabele in de populatie voorkomen.



Scheve verdeling	De hoogste frequenties zitten bij de kleinste of grootste helft van de waarden van de kwantitatieve variabele.
Rechts scheef	De hoogste frequenties zitten bij de grootste helft van de waarden van de kwantitatieve variabele.
Links scheef	De hoogste frequenties zitten bij de kleinste helft van de waarden van de kwantitatieve variabele.
Symmetrische verdeling	De hoogste frequenties zitten rondom de mediaan van de waarden van de kwantitatieve variabele.
Klokvormige verdeling	De frequentieverdeling heeft de symmetrische vorm van een kerstklok.
Tweetoppige verdeling	De hoogste frequenties zitten zowel in de kleinste als in de grootste helft van de waarden van de kwantitatieve variabele. Daartussen zitten waarden met (veel) lagere frequenties.
Meertoppige verdeling	Er zijn meerdere groepjes waarden van de kwantitatieve variabele met een hoge frequentie, met daartussen groepjes met een lage frequentie.
Uniforme verdeling	Iedere waarde van de kwantitatieve variabele heeft dezelfde frequentie als iedere andere waarde.
Staart van de verdeling	Het linker- of rechterdeel van de verdeling heet een staart als er in dat deel heel veel waarden zijn met een heel lage frequentie.
Uitschieter	Staat beschreven bij de boxplot.

Als een kwalitatieve variabele erg veel verschillende waarden kan aannemen, is het verstandig de waarden te groeperen.

Klassen	Groepering van een aantal waarden voor een variabele. Voor continue variabele lengte kun je bijv. klasse 170-<175 maken, waarin de frequentielengtes van 170 tot 175 centimeter worden vastgelegd. Voor discrete variabele leeftijd (in jaren) kun je bijv. klasse 5-9 maken, waarin de frequentieleeftijden 5 tot en met 9 jaar worden vastgelegd.
Klassenbreedte	Lengteklasse 170-<175 heeft een klassenbreedte van 5 centimeter. Leeftijdsklasse 5-9 heeft een klassenbreedte van 5 jaar.
Klassenindeling	De indeling van de waarden van een kwantitatieve variabele in groepen, klassen geheten.
Klassengrenzen	De linker klassengrens van lengteklasse 170-<175 is 170 en de rechtergrens is 175. De linkergrens van leeftijdsklasse 5-9 is 5 en de rechtergrens is 9.
Klassenmidden	Het klassenmidden van lengteklasse 170-<175 is 172,5 centimeter. Het klassenmidden van leeftijdsklasse 5-9 is 7 jaar.

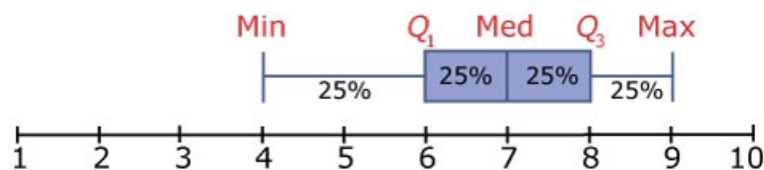
Een dataset kun je samenvatten.

Centrummaat	Getal dat aangeeft waar de waarden omheen zijn gegroepeerd, waar het centrum van de waarden van een variabele zit.
Gemiddelde	De gemiddelde waarde van een kwantitatieve variabele. Het evenwichtspunt van de frequentieverdeling van de variabele.
Mediaan	De middelste waarde van een kwantitatieve variabele, als je alle waarden van klein tot groot op een rij zet.
Modus	De meest voorkomende waarde van de (meestal kwalitatieve) variabele.
Modale klasse, modaal	De klasse of andere groep waarden van de variabele met de hoogste frequentie.

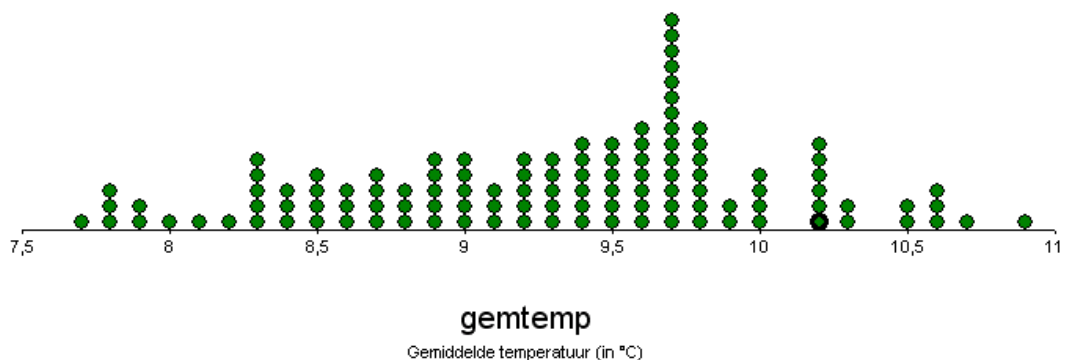
Spreidingsmaat	Getallen die aangeven hoe ver de waarden van een variabele gespreid liggen.
Interkwartielafstand	Q3-Q1.
Q1 (eerste kwartiel)	Rechtergrens van de laagste 25 procent waarden van een kwantitatieve variabele.
Q3 (derde kwartiel)	Linkergrens van de hoogste 25 procent waarden van een kwantitatieve variabele.
Spreidingsbreedte	Maximumwaarde-minimumwaarde van een kwantitatieve variabele.
Standaardafwijking; standaarddeviatie	Maat voor spreiding waarin de afwijking van iedere waarneming ten opzichte van het gemiddelde gewogen wordt.

De waarden van een variabele kun je visualiseren.

Boxplot	Hierin staan de minimumwaarde, Q1, mediaan, Q3 en de maximumwaarde van een kwantitatieve variabele uitgezet boven een getallenlijn. Het deel tussen Q1 en Q3 teken je in de vorm van een doosje (box). Het linker- en rechterkant in de boxplot vormen de snorharen van de boxplot.
Uitschieter	Een waarde die meer dan 1,5 keer de interkwartielafstand onder het eerste kwartiel of boven het derde kwartiel zit.



Dotplot	Een diagram van de waarden van een statistische variabele, met een punt voor ieder element van de populatie.
---------	--



Frequentiepolygoon	Een lijndiagram met de frequenties van de waarden van een variabele.
Cumulatief frequentiepolygoon of somfrequentiepolygoon	Een lijndiagram met de somfrequenties van de waarden van een variabele.
Staafdiagram	Iedere waarde van een variabele heeft een staaf. De hoogte van de staaf wordt bepaald door de frequentie van de waarde.
Steelbladdiagram	Een steelbladdiagram bestaat uit een steel en bladeren. Elk blaadje is – in combinatie met het bijbehorende getal uit de steel – een waarde van de kwantitatieve variabele.

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

Leeftijd van een groep collega's

De waarden van twee kwantitatieve variabelen kun je in samenhang tonen.

Kruistabel	Eigenlijk een gecombineerde frequentietabel. In de bovenste rij van de kruistabel staan de waarden van de ene variabele en in de linkerkolom staan de waarden van de andere variabele. In de cellen staan frequenties.
Spreidingsdiagram	Beide variabelen hebben ieder een eigen as in dit diagram. Ieder element uit de populatie krijgt een eigen punt in dit diagram.
Puntenwolk	De punten in een spreidingsdiagram vormen samen een puntenwolk.



§ 2.1 Data presenteren

§ 2.1.1 Introductie

In het boekje **KIJKEN NAAR DATA** heb je vooral gekeken naar manieren waarop je gegevens tegenkomt in kranten, op internet, e.d. Meestal ontstaan dergelijke statistieken doordat op grond van een bepaalde vraag gericht gegevens zijn verzameld.

Stel dat je wilt weten hoe de lengtes en de gewichten van de 154 leerlingen in havo 4 op een bepaalde school zijn verdeeld. Je **onderzoeksvragen** zouden kunnen zijn:

- Welke conclusies kun je trekken als je de gewichten van de jongens en de meisjes met elkaar vergelijkt?
Dit is een voorbeeld van een vraag waarbij je de eigenschappen van twee groepen vergelijkt.
- Bestaat er een relatie tussen lengte en gewicht? En zo ja, welk?

In deze paragraaf leer je welk soort representaties er zijn om de gegevens uit een onderzoek inzichtelijk te maken.

In paragraaf 2 kijken we naar de verbanden tussen deze representaties.

Vanaf paragraaf 3 gaat het over frequentieverdelingen. We gaan dieper in op frequentieverdelingen en leren deze typeren met kentallen.

In paragraaf 4 vergelijken we twee groepen op basis van de verdeling van één variabele en in paragraaf 5 beschrijven we de samenhang van twee variabelen.

Van de genoemde havo 4-leerlingen zijn behalve hun lengte en hun gewicht nog enkele gegevens opgevraagd. Dat heeft een tabel opgeleverd waarin per leerling de gegevens zijn terug te vinden, zie ► **GEGEVENS154LEERLINGEN**. Die verzameling gegevens noem je een **dataset**.

📋 Opgave 1

Bekijk de genoemde dataset.

- Welke statistische variabelen zijn onderzocht?
- Om de jongens en de meisjes te kunnen vergelijken maak je twee deelgroepen. Je gebruikt dan een van de statistische variabelen als kenmerk om de deelgroepen te onderscheiden. Welke?

§ 2.1.2 Centrale vraag

Wouter is een van de leerlingen in de dataset ► **GEGEVENS154LEERLINGEN**.

Hij is 184 centimeter lang en weegt 68 kilogram. Zijn cijfergemiddelde is 7,5.

Hoe verhoudt Wouter zich tot de rest van de leerlingen in deze dataset?

In deze paragraaf geven we antwoord op deze vraag.

Om dit te doen kijken we naar zeven representaties van de gegevens.

Bij iedere representatie kun je Wouter terugvinden, maar niet elk plaatje geeft dezelfde informatie. We bekijken welke informatie er uit elk van die plaatjes te halen is.

§ 2.1.3 Representaties

In de voorbeelden kijken we steeds naar gewicht. Je gaat daarna zelf kijken naar de informatie over lengte en cijfergemiddelde.

Voor alle opgaven geldt dat je deze zowel in de DWO als in VuStat kunt doen.

Steelbladdiagram

In een steelbladdiagram staan alle leerlingen genoemd. Je kunt precies zien waar Wouter staat.

Jongens	Lengte	Meisjes
	15	666777
14688	16	000011233334555555555555555566677788888888888999
000000123355667888889999	17	00000123333334445555567788
00000000001233333455555567999	18	001125
0000001245	19	6
0	20	

Wouters lengte is 184 centimeter. Voor een jongen is dat ergens boven de helft, er zitten meer jongens onder deze lengte dan erboven. Toch is het niet heel lang, want er zijn nog 22 jongens langer dan Wouter.

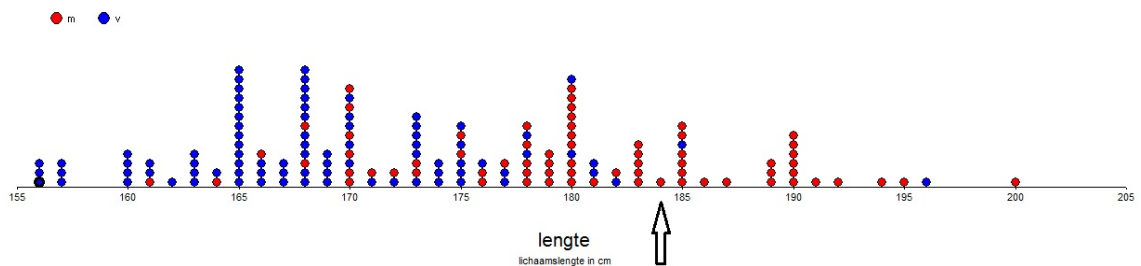
Wanneer Wouter tussen een groep meisjes zou staan, zou hij lang zijn. Want er zijn maar 2 meisjes in deze groep langer dan Wouter. In het totaal zijn dus 24 van de 154 leerlingen langer dan Wouter.

Opgave 2

- Maak zelf een steelbladdiagram voor gewicht en voor cijfergemiddelde. Zorg dat je twee kanten hebt, een voor jongens en een voor meisjes. Wijs Wouter aan in je steelbladdiagrammen.
- Geef een omschrijving van Wouters gewicht en cijfergemiddelde ten opzichte van de rest van de groep.

Dotplot

Ook in een dotplot heeft elke leerling zijn eigen plek.



De plek van Wouter is aangegeven. Het is duidelijk te zien dat Wouter hoort bij de langere leerlingen in de groep.

Het verschil tussen jongens en meisjes wordt aangegeven in kleur. Dit is niet altijd goed te zien, vooral bij een zwart-witprint valt de kleur weg.

Opgave 3

- Maak met gebruik van ICT een dotplot voor *gewicht* en *cijfergemiddelde*.
Wijs Wouter aan in je grafieken.
- Geef een omschrijving van Wouters gewicht en cijfergemiddelde ten opzichte van de rest van de groep. Gebruik alleen de informatie die je in de dotplots kunt zien.

Frequentietabel

In de frequentietabel staat per klasse hoeveel waarnemingen hier in zitten.

Lengte	Freq.
155-159	6
160-164	14
165-169	37
170-174	26
175-179	24
180-184	23
185-189	12
190-194	9
195-199	2
200-204	1
Totaal	154

Wouter zit in de klasse 180-184. Maar je kunt Wouter nu niet meer individueel aanwijzen. Je verliest informatie over het individu wanneer je een frequentietabel met klassen maakt. Toch wordt dit heel vaak gedaan. Uit de tabel blijkt dat 23 leerlingen ongeveer even lang zijn als Wouter. Hij zit niet in de groep waar de meeste leerlingen in zitten, hij is dus langer dan de modale lengte. De lengteklasse van Wouter zit wel net iets boven het gemiddelde. Wouter is dus niet uitzonderlijk lang.

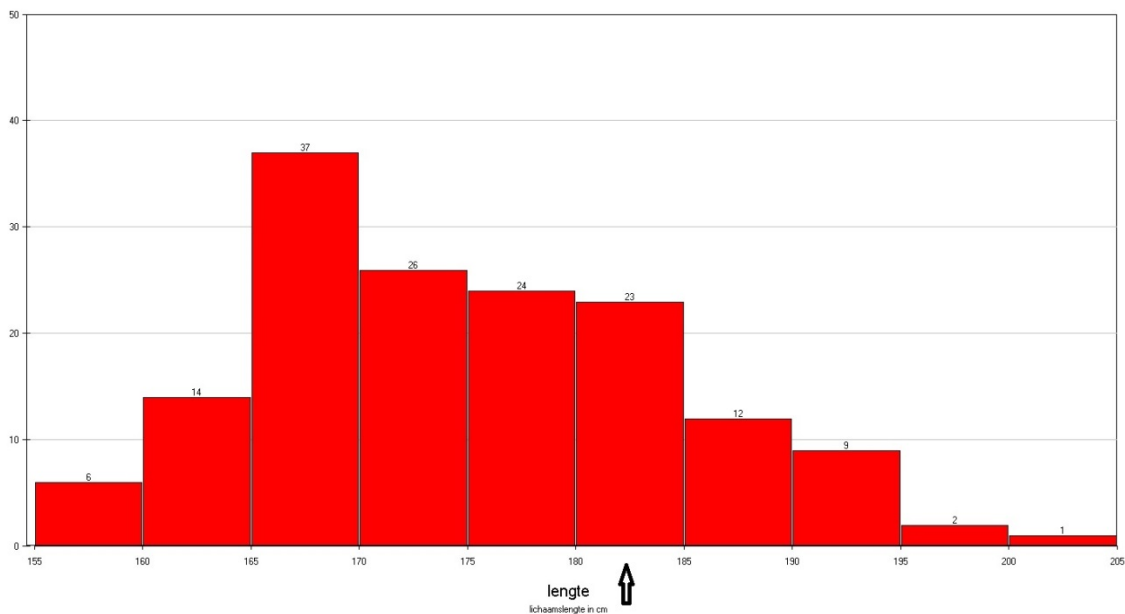
Opgave 4

- Maak met gebruik van ICT een frequentietabel voor gewicht en cijfergemiddelde. Denk na over de klassenindeling (vuistregel: maak ongeveer tien klassen).
- Wat kun je over Wouters gewicht en cijfergemiddelde zeggen wanneer je alleen kijkt naar de frequentietabellen?



Staafdiagram

In een staafdiagram zet je de frequenties uit de frequentietabel in een grafiek.



(Een histogram is een staafdiagram van een kwantitatieve variabele.

De staafjes staan tegen elkaar aan.)

Wouters lengte zit in de klasse 180-185 centimeter. Deze klasse is niet de klasse met de meeste leerlingen, maar er zitten relatief veel leerlingen in deze klasse.

Meer naar rechts in de grafiek wordt het aantal leerlingen in de klassen snel kleiner. Wouter hoort bij de langere leerlingen, maar is niet uitzonderlijk lang.

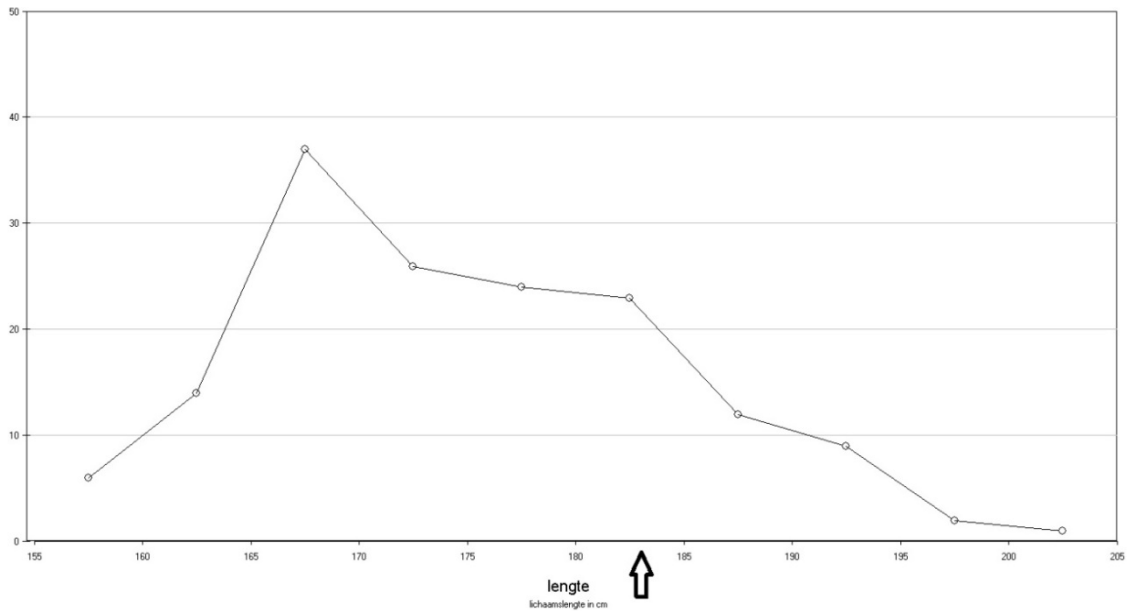
📁 Opgave 5

- Maak een staafdiagram voor gewicht en cijfergemiddelde.
Neem dezelfde klassenindeling als bij opgave 4.
- Wat kun je over Wouters gewicht en cijfergemiddelde zeggen wanneer je alleen naar de staafdiagrammen kijkt?



Frequentiepolygoon

Wanneer je de middens van de bovenkant van het staafdiagram met elkaar verbindt, krijg je een frequentiepolygoon (polygoon = veel hoek). Het frequentiepolygoon geeft dus dezelfde informatie als het staafdiagram.



In het frequentiepolygoon zie je dat de klasse met Wouters lengte aan het einde van de bult zit. Hij hoort dus bij de langere leerlingen, maar is niet uitzonderlijk lang. Het aantal leerlingen dat korter is dan Wouter is duidelijk groter dan het aantal leerlingen dat langer is.

Opgave 6

- Maak voor de variabelen *gewicht* en *cijfergemiddelde* frequentiepolygonen.
- Geef in elk van de grafieken aan waar Wouter zich bevindt en vertel welke informatie over Wouter je uit deze grafieken kunt halen.



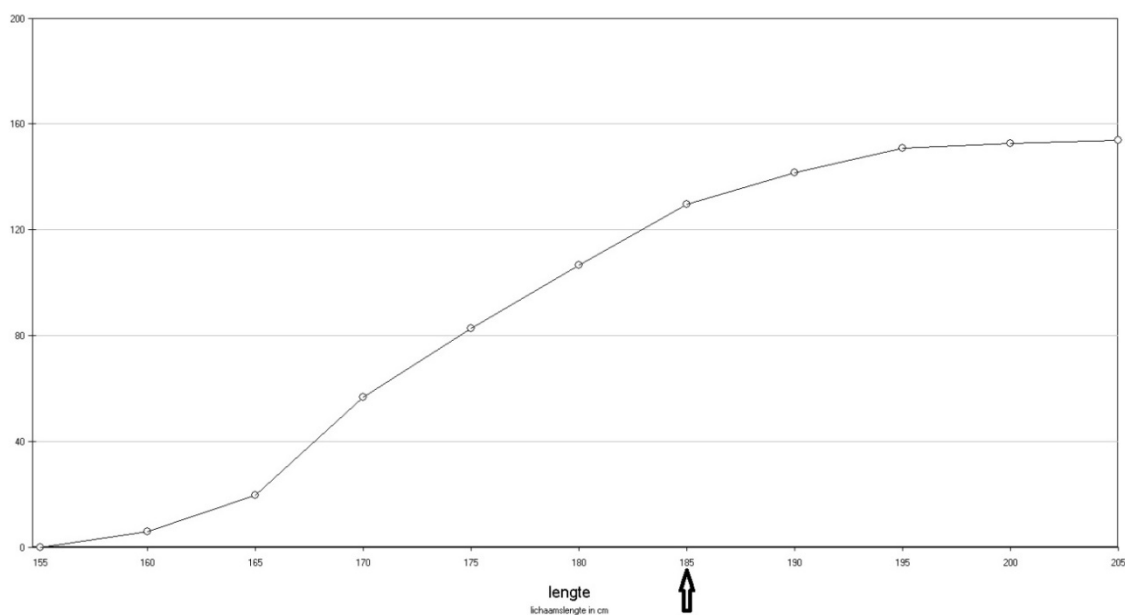
Cumulatief frequentiepolygoon

Lengte	Freq.	Cum. freq.
155-159	6	6
160-164	14	20
165-169	37	57
170-174	26	83
175-179	24	107
180-184	23	130
185-189	12	142
190-194	9	151
195-199	2	153
200-204	1	154
Totaal	154	154

In de frequentietabel is de kolom cumulatieve frequentie toegevoegd. De cumulatieve frequentie is voor elke klasse de som van de frequenties van de voorgaande klassen plus de frequentie van deze klasse.

Bijvoorbeeld, de cumulatieve frequentie van de klasse 170-174 is de som $6 + 14 + 37 + 26 = 83$

Wanneer je deze cumulatieve frequenties in een grafiek zet, krijg je het cumulatieve frequentiepolygoon.



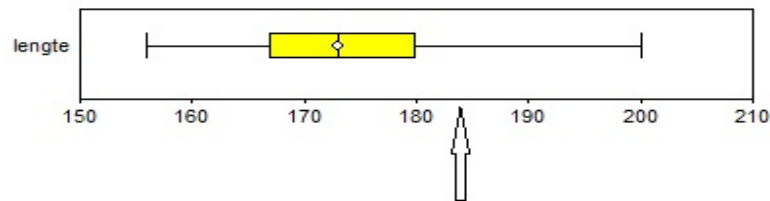
In het cumulatieve frequentiepolygoon kun je aflezen hoeveel leerlingen er langer of minder lang zijn dan de leerlingen in de klasse van Wouter. Je ziet dat na de klasse van Wouter de grafiek nog maar langzaam stijgt, dit betekent dat er niet veel leerlingen langer zijn dan Wouter.

Opdracht 7

- Maak cumulatieve frequentiepolygoon voor de variabelen *gewicht* en *cijfergemiddelde*.
- Wijs in elke grafiek de plek van Wouter aan. Welke informatie over Wouter kun je uit de twee cumulatieve frequentiepolygoon halen?

Boxplot

Bij een boxplot verlies je nog meer individuele gegevens. Om een boxplot te maken gebruik je de mediaan, het eerste en derde kwartiel en het maximum en minimum.



Minimale lengte = 156 centimeter.

Maximale lengte = 200 centimeter.

Mediaan = 173 centimeter.

Q1 = 167 centimeter.

Q3 = 180 centimeter.

Ieder van de vier stukjes van een boxplot bevat 25 procent van de waarnemingen.

Dus:

- Tussen het minimum en Q1 zit 25 procent van de waarnemingen.
- Tussen Q1 en de mediaan zit 25 procent van de waarnemingen.
- Tussen de mediaan en Q3 zit 25 procent van de waarnemingen.
- Tussen Q3 en het maximum zit 25 procent van de waarnemingen.

De plek van Wouter is aangegeven. Wouter zit voorbij de box van de boxplot, dit betekent dat hij hoort bij de langste 25 procent van de leerlingen. Het is niet duidelijk hoe de verdeling binnen deze groep is, je kunt alleen nog zeggen dat hij niet het langste is.

📊 Opgave 8

- a. Maak boxplotten voor de variabelen *gewicht* en *cijfergemiddelde*.

Geef de plek van Wouter aan in beide boxplotten.

Wat kun je zeggen over Wouter ten opzichte van de rest van de leerlingen?

Antwoord op de centrale vraag

Uit alle representaties van de data die in deze paragraaf zijn behandeld blijkt dat Wouter een redelijk lange jongen is. Hij is niet uitzonderlijk lang, maar behoort tot de 25 procent langste leerlingen in dit onderzoek. Er zijn 24 leerlingen ongeveer even lang als hij, dat zijn 22 jongens en 2 meisjes.

§ 2.1.4 Oefenen

■ Opgave 9

Bekijk de genoemde dataset ► **GEGEVENS154LEERLINGEN**.

- Hoe lang is het grootste meisje? En de grootste jongen?
- Welke lengtes komen het meeste voor?
- Is het berekenen van gemiddelden een goede manier om de lengtes van de meisjes en de jongens met elkaar te vergelijken? Licht je antwoord toe.

■ Opgave 10

- Maak een dotplot van de lengtes van de leerlingen. Maak geen onderscheid tussen jongens en meisjes, je krijgt dus één dotplot met alle leerlingen erin.
- Beantwoord nu de vragen a en b van opgave 9 met behulp van deze figuur.
- Wat is het voordeel van een dotplot ten opzichte van de lijst met gegevens die je in de dataset hebt?
- Je ziet dat de lengtes 170, 180 en 190 opvallend vaak voorkomen. Heb je daar een verklaring voor?
- In plaats van een dotplot had je ook een steelbladdiagram kunnen maken, daar kun je dezelfde informatie uithalen.
Geef een voor- en nadeel van het gebruik van een steelbladdiagram ten opzichte van het gebruik van een dotplot.

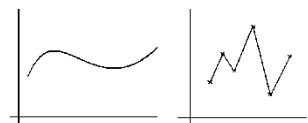
Opgave 11

Je hebt al kennism gemaakt met kwalitatieve en kwantitatieve statistische variabelen.

- Noem van beide soorten variabelen een voorbeeld.
- Aan de variabele geslacht worden soms twee waarden toegekend: 0 = vrouw en 1 = man. Wordt de variabele daarmee kwantitatief?
- De lengtes bij een bevolkingsonderzoek worden gemeten in centimeters. Kun je daarvoor redenen aangeven?
- Je ziet hier twee weegschalen. Wat is het verschil tussen beide als het gaat om het aflezen van een gewicht?



- Bij een grafiek van het temperatuurverloop van een dag kun je een vloeiende lijn tekenen. Waarom kan dat niet bij een grafiek van de gemiddelde maandtemperatuur in 2009?



Opgave 12

Ga van de volgende statistische variabelen na of ze continu of discreet zijn en welke waarden ze kunnen aannemen.

- Geboortejaar (van nog levende personen).
- Temperatuur op de Noordpool in graden Celsius.
- Enquête met een driepuntsschaal.
- Gewicht van muizen in grammen.
- Toetscijfer.
- Profiel in bovenbouw havo.
- Kwaliteit van een hotel: aantal sterren.

📊 Opgave 13

Bij het maken van frequentietabellen en staafdiagrammen bij de lengtes en de gewichten van de 154 leerlingen speelt de volgorde waarin je de gemeten waarden zet een grote rol.

- Bekijk de variabele *profielkeuze*. Kun je daarbij een zinvol staafdiagram maken?
En is de volgorde van de staven dan van belang?
- Bekijk de variabele *huiswerk*. Kun je daarbij een zinvol staafdiagram maken?
En is de volgorde van de staven dan van belang?
- Bekijk de variabele *geboortejaar*.
Waarom is het bij deze dataset nauwelijks zinvol om hierbij een frequentietabel te maken?
- Bekijk de variabele *plezier*. Kun je daarbij een zinvol staafdiagram maken?
En is de volgorde van de staven dan van belang?
- Hier wordt een aantal andere statistische variabelen genoemd. Beschrijf bij elk van deze variabelen of deze kwalitatief of kwantitatief is.
Geef bij kwantitatieve variabelen aan of het een continue of discrete variabele is.
 - Dagelijkse reistijd naar school* (in minuten).
 - Aantal lesuren per week*.
 - Muziekvoorkeur* (bedenk zelf de categorieën).
 - Belang van bewegingsonderwijs voor iedereen*
(vijfpuntsschaal: totaal onbelangrijk, niet erg belangrijk, neutraal, belangrijk, heel belangrijk).

Opgave 14

Iemand wil een onderzoek doen bij examenklassen havo met de volgende variabelen: *geslacht*, *geboortejaar*, *geboortemaand*, *gewicht*, *lengte*, *cijfergemiddelde*, *cijfer voor wiskunde*, *huiswerk*, *wiskundegroep*, *profiel*, *plezier*.

Geef bij elk van deze variabelen aan of deze kwalitatief of kwantitatief is en welke waarden deze kan aannemen.

📊 Opgave 15

- Als je zonder te kijken in de gegevens iets zou moeten zeggen over het verschil in lengte tussen jongens en meisjes, wat zou je dan zeggen?
- Maak twee dotplots voor lengte. Eén voor jongens en één voor meisjes.
Vergelijk de lengtes van meisjes en jongens. Wat valt je op?
Kun je het antwoord dat je bij opgave a bedacht hebt terugvinden?

Opgave 16

Je kunt de lengtes van de jongens en de meisjes ook in een staafdiagram zetten.

Gebruik nu niet de absolute aantallen, maar de relatieve aantallen (procenten).

- Maak twee staafdiagrammen van de lengtes: één voor jongens en één voor meisjes van de relatieve frequenties.
- Waarom kan het nuttig zijn om frequenties om te zetten naar relatieve frequenties wanneer je twee groepen wilt vergelijken?
- Hoeveel procent van de jongens is langer dan 180 centimeter? En hoeveel procent van de meisjes?
- Bekijk de 50 procent kleinste meisjes.
Tussen welke waarden zit hun lengte? En hoe zit dat bij de jongens?
- Bekijk de 25 procent grootste meisjes.
Tussen welke waarden zit hun lengte? En hoe zit dat bij de jongens?
- Vergelijk de grafiek van de jongens met die van de meisjes. Wat valt je op?

Opgave 17

Bij het maken van een frequentietabel deel je de gegevens in in klassen. Je krijgt een ander beeld wanneer je andere klassen neemt. De klassen moeten wel altijd even groot zijn, maar je kunt ze groter of minder groot maken.

- Maak vier staafdiagrammen van de lengte.
Maak de klassenbreedte eerst 2, dan 5, dan 10 en tot slot 15 centimeter breed.
- Wat valt je op aan de verschillende staafdiagrammen?
Welke klassenindeling vind jij hier het beste passen en waarom?
- Je wilt weten hoeveel procent van de meisjes tussen de 155 en 160 centimeter lang is.
Welke tabel of grafiek heb je nodig om deze vraag te kunnen beantwoorden?
Beantwoord de vraag.
- Hoeveel procent van de jongens is langer dan 182 centimeter?
Met welke representatie heb je deze vraag beantwoord?
- Welke voordelen heeft het groeperen van de metingen in klassen?
- Welke nadelen heeft het groeperen van de metingen in klassen?

Opgave 18

Je kunt een klassenindeling op verschillende manieren noteren.

- Lengtes van de bladeren van een bepaalde soort boom (in cm) worden ingedeeld in de volgende klassen: 6,5-<7,5; 7,5-<8,5; enz.
Bepaal de klassenbreedte en de klassenmiddens.
- De leeftijden van de werknemers van een bepaald bedrijf worden ingedeeld in de volgende klassen: 20-24, 25-29, ..., 60-64.
Bepaal de klassenbreedte en de klassenmiddens.
- Een theater houdt bij hoeveel kaartjes er voor een voorstelling worden verkocht.
De klasse 200-249 geeft het aantal voorstellingen weer waarvoor 200 tot en met 249 kaartjes verkocht zijn. Bepaal de klassenbreedte en het klassenmidden van deze klasse.
- Bij welke variabelen uit de dataset met gegevens van 154 leerlingen is het zinvol en/of mogelijk om een klassenindeling te maken? Licht je antwoorden toe.



Opgave 19

We kijken nu naar de verschillen tussen jongens en meisjes in het aantal uren dat ze besteden aan huiswerk.

- Maak twee frequentietabellen van het aantal uren dat aan huiswerk wordt besteed. Eén voor meisjes en één voor jongens. Voeg een kolom relatieve frequenties toe (klassenbreedte = 5).
- Waarom is het verstandig om hier te kijken naar relatieve frequenties?
- Je ziet dat de huiswerkuren in klassen zijn ingedeeld. Noem een voor- en nadeel van het gebruik van klassen.
- Teken van beide frequentietabellen de bijbehorende relatieve frequentiepolygonen.
- Wat kun je zeggen over het verschil tussen meisjes en jongens op basis van de tabel en/of de frequentiepolygonen?
- Wat vind je overzichtelijker, de tabel of de polygonen?

Opgave 20

Voor een practicum biologie zijn op twee velden regenwormen gevangen. Vervolgens werden de lengtes van die wormen gemeten. In de tabel rechtsonder zie je de resultaten.

- Om wat voor soort statistische variabele gaat het hier?
- Wat is de maximaal mogelijke lengte van een gevangen regenworm?
- Wanneer is het verstandig om relatieve frequenties te gebruiken in plaats van absolute?
- Maak van beide velden relatieve frequentietabellen van de lengtes en teken er staafdiagrammen bij.
Dit mag je ook doen met ICT. Je moet dan wel de gegevens zelf invoeren.
- Vergelijk beide staafdiagrammen. Wat valt je daarbij op?

lengte regenworm (in cm)	aantal veld 1	aantal veld 2
0 – 2	4	1
3 – 5	7	1
6 – 8	15	5
9 – 11	20	9
12 – 14	23	15
15 – 17	15	18
18 – 20	4	12
21 – 23	1	6
24 – 26	1	3

Opgave 21

Voor een bepaalde toets kun je maximaal 100 punten scoren. Hier zie je hoe een groep van 40 personen de toets heeft gemaakt.

59 57 53 60 63 58 77 33 50 59 58 75 62 54 53 78 59 68 65 62
57 60 80 47 90 30 60 35 57 87 63 65 63 58 65 70 73 58 63 55

- Om wat voor soort statistische variabele gaat het hier?
- Deel deze scores in klassen in, neem als laagste klasse 25-<35. Maak een frequentietabel.
- Maak bij deze tabel een staafdiagram van relatieve frequenties.
- Personen die 55 of meer punten hebben behaald, scoren voldoende.
Hoeveel procent van deze groep scoorde voldoende?
- Je had ook als eerste klasse 30-<40 kunnen nemen. Wat is daarvan het nadeel?



Opgave 22

Er zijn een paar belangrijke dingen waar je op moet letten bij het tekenen van een cumulatief frequentiepolygoon:

- De eerste stip begint aan de linkerkant van de kleinste klasse.
 - De volgende stippen staan steeds aan de rechterkant van de klasse.
- a. Leg uit waarom de twee hierboven genoemde eigenschappen van een cumulatief frequentiepolygoon heel logisch zijn.
- b. Bekijk de cumulatieve frequentiepolygoon van opgave 7.
Wordt hier aan beide voorwaarden voldaan?

📊 Opgave 23

In deze opgave komt Wouter uit de centrale vraag weer even terug en samen met hem ook weer dataset ► **GEGEVENS154LEERLINGEN**.

- a. Maak met behulp van ICT cumulatieve frequentiepolygoon van de lengte, het gewicht en het cijfergemiddelde van de leerlingen, met langs de zijkant percentages i.p.v. de absolute frequenties.
- b. Geef een voor- en nadeel van het gebruik van relatieve frequenties ten opzichte van het gebruik van absolute frequenties.
- c. Vul de volgende zin aan: Wouter behoort bij de ... procent langste leerlingen.
- d. Maak ook een dergelijke zin voor Wouter over gewicht en cijfergemiddelde.
- e. Wanneer je uitspraken doet zoals hierboven bij opgave c en d, doe je eigenlijk een aanname over de verdeling van de mensen in een klasse.
Welke aanname is dat? Vind je dit een gerechtvaardigde aanname?

📊 Opgave 24

- a. Splits de data in dataset ► **GEGEVENS154LEERLINGEN** in jongens en meisjes en maak nu drie boxplots:
- Lengte.
 - Uren huiswerk.
 - Cijfergemiddelde.
- b. Geef het verschil tussen jongens en meisjes op basis van deze drie boxplots.

📊 Opgave 25

- a. Maak nogmaals het relatieve cumulatieve frequentiepolygoon voor de lengte.
- b. Print deze grafiek en geef in de grafiek de mediaan aan.
- c. Geef in de grafiek het eerste kwartiel en derde kwartiel aan.
- d. Teken onder het relatieve cumulatieve frequentiepolygoon met de hand de boxplot.
- e. Omschrijf hoe je uit een relatief cumulatief frequentiepolygoon de gegevens kunt halen om een boxplot te tekenen.



§ 2.1.5 Om te onthouden

Ruwe data orden je in tabellen of diagrammen waarin je de **frequenties** uitzet tegen de waarden die de statistische variabele kan aannemen. Zo krijg je een **frequentieverdeling** van de gegevens.

Kwantitatieve variabelen kun je onderscheiden in:

- **Continue variabelen**: alle waarden binnen een interval kunnen worden aangenomen.
- **Discrete variabelen**: alleen bepaalde waarden kunnen worden aangenomen.

Vaak verdelen we de gegevens in **deelgroepen**, bijvoorbeeld jongens/meisjes.

Dan kun je iets zeggen over de verschillen tussen deze deelgroepen.

- Door de ruwe data in **klassen** te groeperen krijg je een beter overzicht, maar dan zijn de ruwe data zelf niet meer te zien.
- Met indelen in klassen kun je gegevens groeperen. De lengtes kun je bijvoorbeeld groeperen in klassen als 160-<165, enzovoorts. De klassen hebben dan een **klassenbreedte** van 5 centimeter. Het **klassenmidden** is 162,5 en de getallen 160 en 165 zijn de **klassengrenzen**.



§ 2.1.6 Geïntegreerd oefenen

Opgave 26

In 1947 hielden de wiskundigen Freudenthal en Sittig een statistisch onderzoek ten behoeve van een nieuw maatsysteem voor vrouwenkleding in opdracht van het warenhuis De Bijenkorf. Onder andere maten zij de mouwlengte van 5001 vrouwen in centimeters nauwkeurig.

Hiernaast zie je een frequentietabel met hun data.

- Met wat voor soort variabele heb je hier te maken?
- Maak een klassenindeling: 45-49, 50-54, enz. Maak bij die klassenindeling een staafdiagram van relatieve frequenties.
- Vergelijk deze klassenindeling met de gegeven frequentietabel en beschrijf het voor- en nadeel ervan.
- Hoeveel procent van deze vrouwen heeft een mouwlengte van 53 centimeter?
- Hoeveel procent van deze vrouwen heeft een mouwlengte van 65 centimeter of meer?

mouw- lengte	frequentie
49	3
50	11
51	22
52	53
53	89
54	163
55	250
56	405
57	519
58	660
59	578
60	653
61	560
62	421
63	260
64	159
65	106
66	52
67	18
68	15
69	3
70	0
71	1
	5001

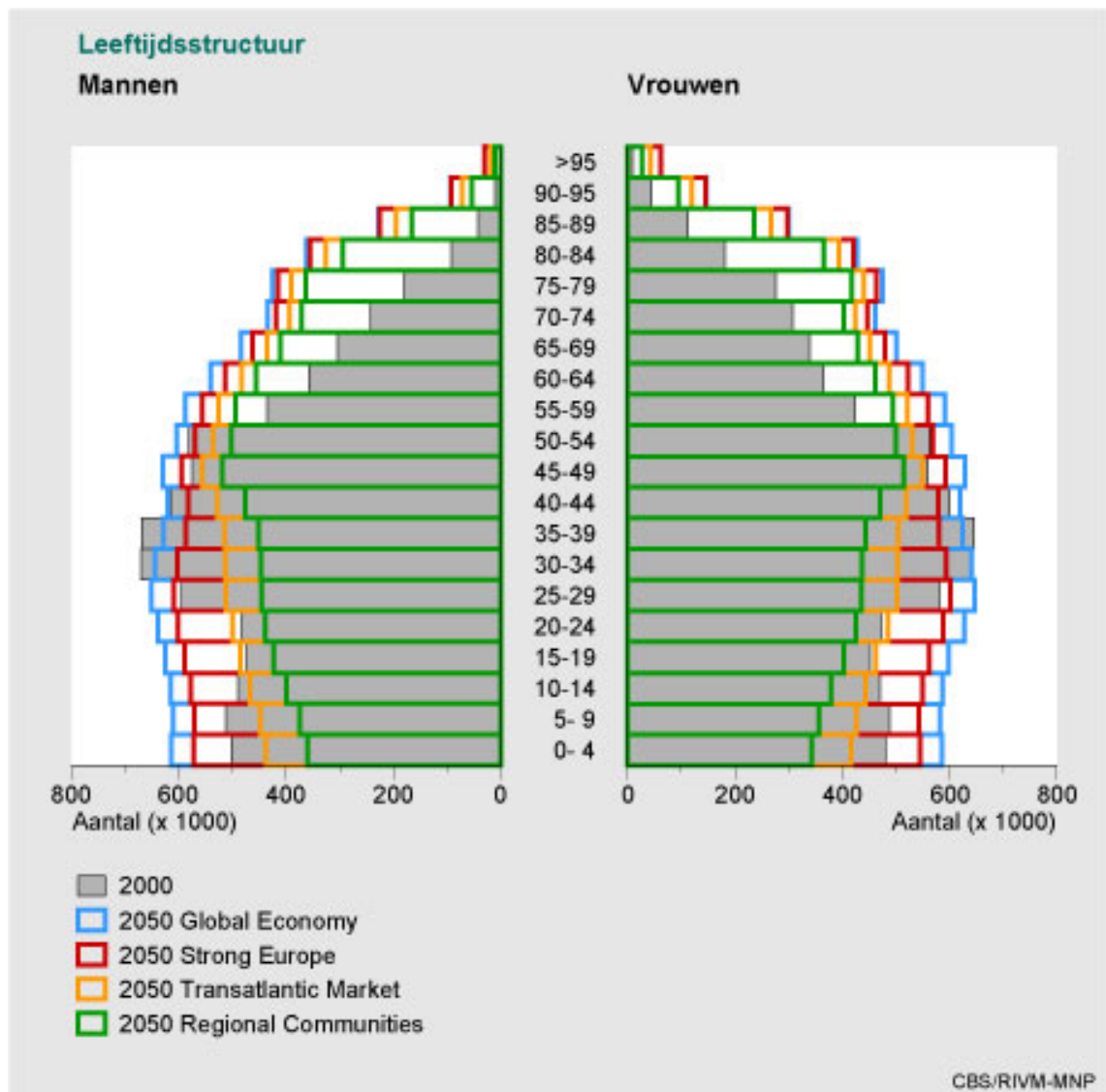
📖 Opgave 27

Op de volgende pagina zie je het leeftijdsdiagram voor Nederland in het jaar 2000.

Verder zijn er vier verschillende prognoses gedaan voor 2050.

- Bepaal de klassenbreedte en het klassenmidden van de eerste klasse.
- Hebben alle klassen dezelfde breedte?
- Waarom staan in dit leeftijdsdiagram absolute frequenties en geen relatieve frequenties?
- Je kunt dit leeftijdsdiagram omzetten naar relatieve frequenties door percentages van het totaal aantal Nederlanders te nemen of door percentages van de aantallen mannen en vrouwen afzonderlijk te nemen. Noem van elke mogelijkheid een voordeel en licht je antwoord toe.
- Hoe kun je zien dat vrouwen gemiddeld langer leven dan mannen?
- De vier prognoses zijn gebaseerd op vier economische scenario's. Bij welk scenario is het vergrijzingsprobleem het sterkst in Nederland?





📁 Opgave 28

Gebruik het bestand ► **SPORTPRESTATIES**. Je vindt er gegevens van brugklassers op sportgebied.

- Welke vijf statistische variabelen tref je aan in deze dataset?
Meld bij elke variabele om welke soort het gaat.
- Bij het vergooien gaat het om de geworpen afstand in meter met een kogel van 200 gram.
Kies hierbij een geschikte klassenindeling en maak frequentietabellen en staafdiagrammen voor de jongens en de meisjes afzonderlijk.
- Wat valt je op wat het vergooien betreft?

§ 2.2 Verbanden tussen datarepresentaties

§ 2.2.1 Introductie

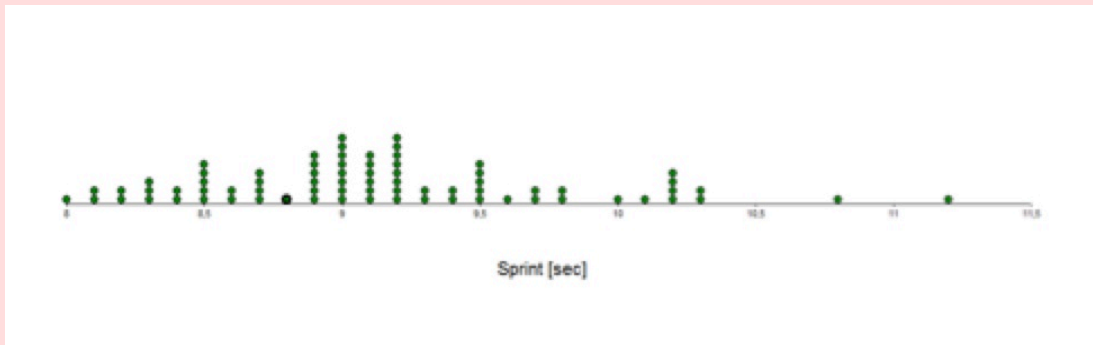
In paragraaf 1 heb je een hele reeks aan datarepresentaties leren kennen.

In deze paragraaf leer je welke verbanden er tussen deze representaties bestaan.

Ook leer je van iedere representatie welke informatie je er bij uitstek goed uit kunt aflezen en wanneer je welke representatie het beste kunt gebruiken.

§ 2.2.2 Centrale vraag

Je ziet hier de dotplot van de sprinttijden van een groep brugklasleerlingen.



Centrale vraag 1

Kun je vaststellen hoe het staafdiagram of de boxplot van deze sprinttijden eruit ziet?

Centrale vraag 2

Welke informatie kun je over deze sprinttijden aflezen uit de dotplot?

En welke informatie kun je beter aflezen uit een andere representatie?



§ 2.2.3 Verbanden tussen representaties

Van dotplot naar andere representaties

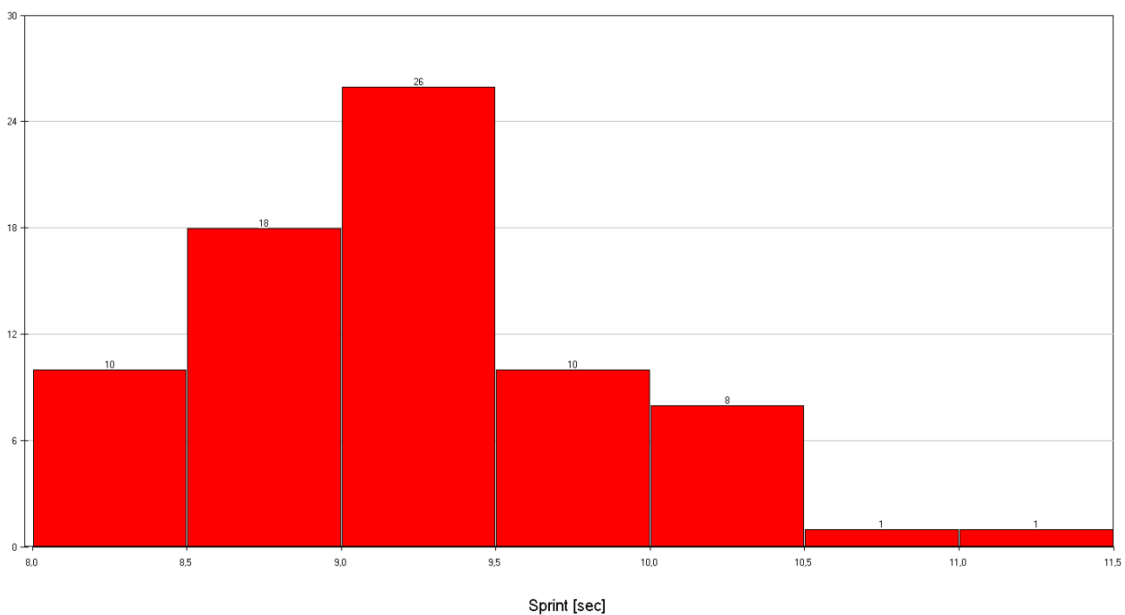
Opgave 29

- Schets – op basis van de dotplot uit de centrale vraag – het staafdiagram en het frequentiepolygoon voor de sprinttijden.
- Schets ook het steelbladdiagram, zonder precies de steel en de blaadjes allemaal in te vullen: het gaat om de vorm van de blaadjes.
- Omschrijf het verband tussen de dotplot, het staafdiagram, het frequentiepolygoon en de vorm van het steelbladdiagram.

Opgave 30

Je kunt de sprinttijden ook indelen in klassen.

Je ziet hier het staafdiagram (histogram) van zo'n klassenindeling van de sprinttijden.



- Had je de vorm van dit staafdiagram kunnen voorspellen vanuit de dotplot uit de centrale vraag? Beargumenteer je antwoord.
- Had je de vorm van de dotplot uit de centrale vraag kunnen voorspellen op basis van dit staafdiagram? Beargumenteer je antwoord.



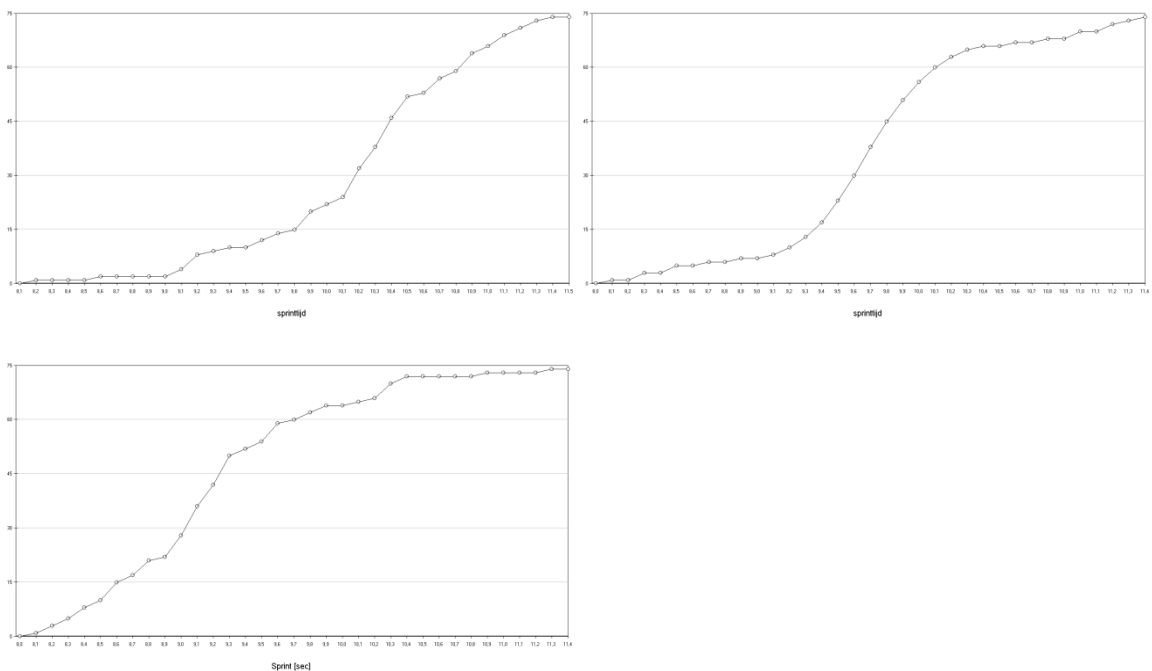
Opgave 31

Kun je voorspellen hoe de boxplot eruit ziet die bij de dotplot van de sprinttijden hoort?

Splits de beantwoording van deze vraag uit in de volgende deelvragen en beargumenteer ieder antwoord, gebruik makend van de gegevens uit de dotplot:

- Denk je dat de linkerhelft langer/korter/even lang is als de rechterhelft van de totale boxplot?
- Denk je dat de kleinste 25 procent langer/korter/even lang is als de grootste 25 procent van de boxplot?
- Verwacht je op basis van de dotplot (dus zonder te rekenen) mogelijke uitschieters in de dataset?

Opgave 32



Welke van deze drie cumulatieve frequentiepolygonen hoort bij de dotplot uit de centrale vraag? Beargumenteer je antwoord en gebruik daarbij de gegevens uit de dotplot.

Antwoord op centrale vraag 1

Het is eenvoudig om vanuit een dotplot het bijbehorende staafdiagram te maken: je ziet bij wijze van spreken de staafjes al staan. Voor een dotplot, een staafdiagram en een frequentiepolygoon geldt: als je de één hebt, heb je eigenlijk ook meteen de andere twee diagrammen.

Vanuit een staafdiagram kun je ook eenvoudig een dotplot, een staafdiagram en een frequentiepolygoon afleiden: andersom geldt dat alleen voor de vorm van het staafdiagram.

Om vanuit een dotplot te kunnen voorspellen hoe een diagram met klassenindeling, een cumulatieve frequentiepolygoon of een boxplot eruit zien is moeilijker, maar het is wel degelijk mogelijk.

Verschillende representaties, verschillende informatie

Opgave 33

Hieronder zie je de frequentietabel met de sprinttijden uit de dotplot van de centrale vraag.

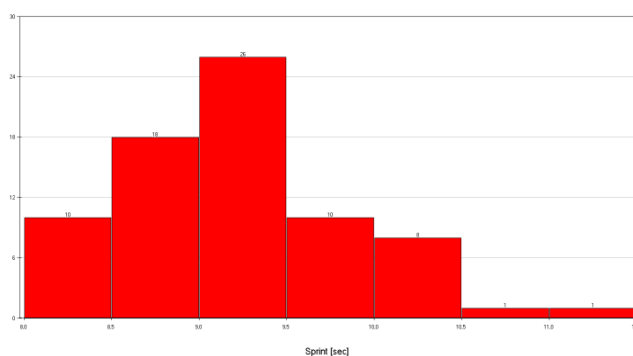
Sprint [sec]	Freq.	Sprint [sec]	Freq.	Sprint [sec]	Freq.	Sprint [sec]	Freq.
8,0	1	9,0	8	10,0	1	11,0	0
8,1	2	9,1	6	10,1	1	11,1	0
8,2	2	9,2	8	10,2	4	11,2	1
8,3	3	9,3	2	10,3	2	11,3	0
8,4	2	9,4	2	10,4	0		
8,5	5	9,5	5	10,5	0		
8,6	2	9,6	1	10,6	0		
8,7	4	9,7	2	10,7	0		
8,8	1	9,8	2	10,8	1		
8,9	6	9,9	0	10,9	0		
						Totaal	74

Krijg je meer, minder of dezelfde informatie over de sprinttijden als je alleen de frequentietabel hebt in vergelijking met de dotplot? Beargumenteer je antwoord.

Opgave 34

Je ziet hier het staafdiagram en de frequentietabel van de sprinttijden, beide ingedeeld in klassen.

Sprint [sec]	Freq.
8,0-8,4	10
8,5-8,9	18
9,0-9,4	26
9,5-9,9	10
10,0-10,4	8
10,5-10,9	1
11,0-11,4	1
Totaal	74



- Noem de overeenkomsten tussen het staafdiagram en de frequentietabel.
- Stel dat je alleen deze frequentietabel met klassenindeling van de sprinttijden hebt. Hoe zou dan het staafdiagram met een klassenbreedte van 0,1 seconde eruit komen te zien? Maak er een schets van en leg uit waarom deze schets afwijkt van het staafdiagram dat je in opgave 29 geschetst hebt.

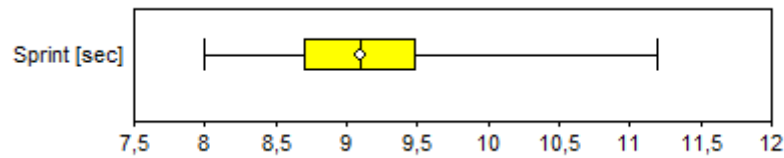
Opgave 35

Bekijk nogmaals de frequentietabel en het staafdiagram in de vorige opgave.

Kun je er makkelijk uit aflezen wat de meest gesprinte tijden waren? Kun je dat in de dotplot uit de centrale vraag net zo makkelijk of misschien juist moeilijker aflezen? En uit de frequentietabel uit opgave 33? Waarom wel of niet?

Opgave 36

Je ziet hier de boxplot die bij de dotplot uit de centrale vraag hoort.



- In welk tijdsinterval liepen de 25 procent langzaamste leerlingen hun sprint?
- Hoe snel liep de allerlangzaamste leerling?
- Kun je deze twee gegevens ook zo direct aflezen uit de dotplot?

Antwoord op centrale vraag 2

Uit een dotplot kun je precies aflezen welke andere sprinttijden er zijn gelopen. Je krijgt een grof inzicht over hoe snel of hoe langzaam een individuele leerling was ten opzichte van de rest van de klas. Andere representaties kunnen dit laatste beter: in een boxplot kun je bijvoorbeeld direct zien hoe snel de langzaamste 25 procent gelopen heeft.



§ 2.2.4 Oefenen

Opgave 37

Esmee, een van de brugklasleerlingen, deed 8,5 seconden over de sprint.

- a. Geef per representatie aan of je er Esmee's sprinttijd direct uit kunt aflezen en zo ja, hoe?
Zo nee: waarom niet?

Representatie	Esmee's tijd is direct af te lezen	
	Ja/nee	Zo ja: hoe? Zo nee: waarom niet?
Dotplot		
Frequentietabel		
Staafdiagram (histogram)		
Frequentiepolygoon		
Bovenstaande diagrammen met klassenindeling		
Steelbladdiagram		
Cumulatief frequentiepolygoon		
Boxplot		

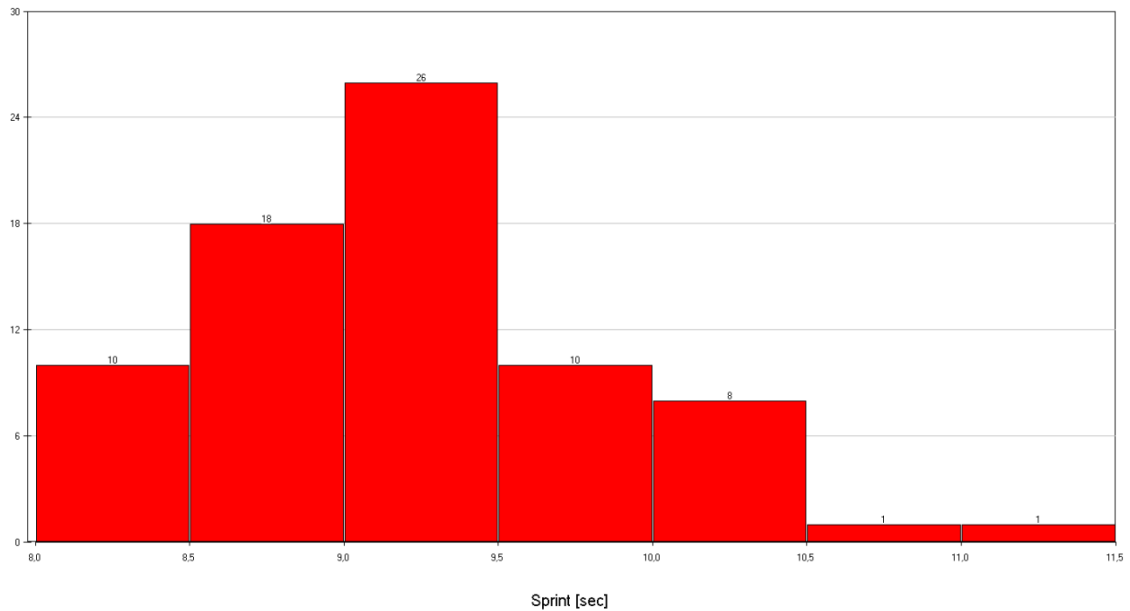
- b. Geef per representatie aan of je er direct (dus zonder tellen of rekenen) uit kunt aflezen hoeveel klasgenoten sneller waren dan Esmee en zo ja, hoe?

Representatie	Direct afleesbaar hoeveel klasgenoten sneller zijn dan Esmee	
	Ja/nee	Zo ja: hoe? Zo nee: waarom niet?
Dotplot		
Frequentietabel		
Staafdiagram (histogram)		
Frequentiepolygoon		
Bovenstaande diagrammen met klassenindeling		
Steelbladdiagram		
Cumulatief frequentiepolygoon		
Boxplot		



Opgave 38

Je ziet hier het staafdiagram van de sprinttijden, ingedeeld in klassen van 5 seconden.



Is het mogelijk om vanuit dit staafdiagram een zinvol steelbladdiagram te maken?

Zo ja: maak het steelbladdiagram; zo nee: beargumenteer je antwoord.

Opgave 39

Bekijk de dataset ► **GEGEVENS154LEERLINGEN**.

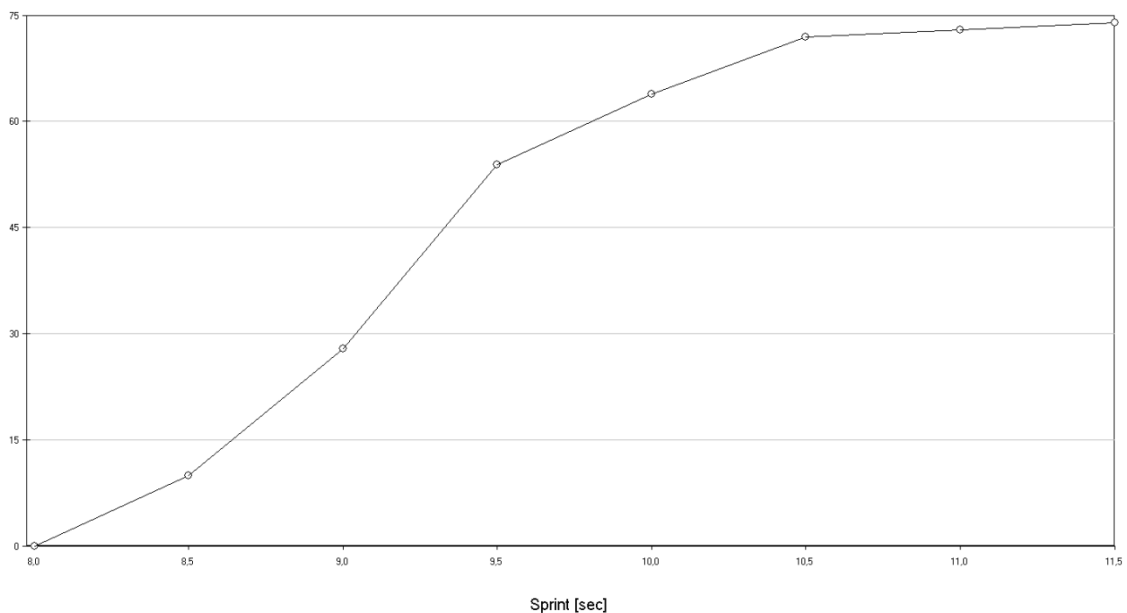
- Maak een frequentietabel en een steelbladdiagram van de variabele huiswerk.
- Bedenk tenminste twee redenen waarom een frequentietabel soms handiger is dan een steelbladdiagram.
Is het in deze situaties erg dat je de gegevens van een individuele waarneming mist?
- Maak een staafdiagram van de variabele huiswerk, neem klassenbreedte 2.
Hoeveel klassen heb je nu nodig?
- In welk soort situaties zou jij gebruik maken van een frequentietabel?
En wanneer van een staafdiagram? Bedenk voor beide een voor- en nadeel.



Opgave 40

In de vorige paragraaf heb je ook geleerd om een boxplot bij een (relatieve) cumulatieve frequentiepolygoon te tekenen.

- a. Teken in het onderstaande cumulatieve frequentiepolygoon van de sprinttijden de horizontale lijnen die horen bij 25, 50 en 75 procent en teken er de boxplot onder.



- b. Vergelijk beide diagrammen met elkaar. Is er informatie die je wel uit het cumulatieve frequentiepolygoon kunt halen maar niet uit de boxplot, of andersom? Beargumenteer je antwoord.

§ 2.2.5 Om te onthouden

Verschillende diagrammen geven verschillende informatie.

Als we weer even het voorbeeld nemen van de sprinttijden:

- Een dotplot en een staafdiagram tonen iedere sprinttijd afzonderlijk.
- Een frequentietabel, een staafdiagram en een frequentiepolygoon tonen de sprinttijd samengevoegd met andere sprinttijden, al naar gelang de gebruikte klassenindeling. Hoe groter de klassen zijn, hoe minder je van ieder sprinttijd terug ziet, maar hoe meer je kunt zeggen over groepen sprinttijden (de meest gesprinte tijden, het tijdsinterval van de langzaamste sprinters etc.).
- Een cumulatief frequentiepolygoon en een boxplot geven vooral informatie over groepen sprinttijden, net als een indeling in grote klassen. In een relatief cumulatief frequentiepolygoon en in een boxplot kun je meteen eigenschappen zien van bijv. de snelste 25 procent sprinttijden, de 50 procent langzaamste sprinttijden etc.

Het heeft dus zin om na te denken over welk diagram je van je dataset wilt maken: afhankelijk van je toepassing kies je het meest geschikte diagram.

§ 2.2.6 Geïntegreerd oefenen

Opgave 41

Een bedrijf heeft 25 werknemers in vaste dienst met een volledige werkweek. De netto weeklonen van deze werknemers zijn verwerkt in de frequentietabel rechtsonder. De weeklonen zijn verdeeld in klassen met een breedte van 50. De ruwe data zijn niet bekend.

- Waarom kun je vanuit deze frequentietabel de gemiddelde weeklonen niet meer precies uitrekenen, maar alleen nog schatten?
- Bepaal de klassenmiddens en bereken hiermee dit geschatte gemiddelde.
- Waarom kun je vanuit een klassenindeling zoals deze niet meer een nauwkeurige boxplot maken?

netto weekloon (in €)	aantal werknemers
400 – < 450	2
450 – < 500	3
500 – < 550	4
550 – < 600	8
600 – < 650	3
650 – < 700	2
700 – < 750	2
750 – < 800	1

Opgave 42

Gebruik het bestand ► **SPORTPRESTATIES**. Je vindt er gegevens van brugklassers op sportgebied. Bekijk dit keer de gegevens over het verspringen.

Tjeerd – een van de brugklassers – sprong 330 centimeter ver.

Met welke representatie(s) kun je het beste laten zien hoe goed hij eigenlijk gesprongen heeft ten opzichte van de rest van klas? En welke representatie(s) zullen Tjeerd het minst bevallen?

Maak ze en beargumenteer waarom ze goed of juist niet goed gebruikt kunnen worden

(vergeet niet dat je ook verschillende klassenindelingen kunt gebruiken!).

Opgave 43

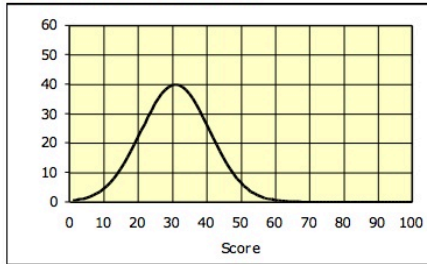
Je ziet op de volgende twee pagina's (en in de DWO) acht frequentiepolygonen (A tot en met H) en acht boxplots (1 tot en met 8). Voor elk van de frequentiepolygonen is er een bijbehorend boxplot.

- Leg of sleep de bij elkaar horende duo's naast elkaar.
- Check: zijn dit dezelfde duo's als je buurman/buurvrouw heeft samengesteld?
- Leg aan je buurman/buurvrouw per duo uit waarom je vindt dat ze bij elkaar horen en maak samen een definitieve duosamenstelling.

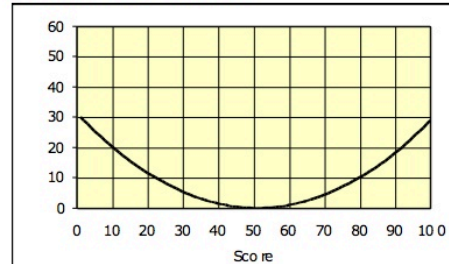
(© 2012 MARS, Shell Center, University of Nottingham)



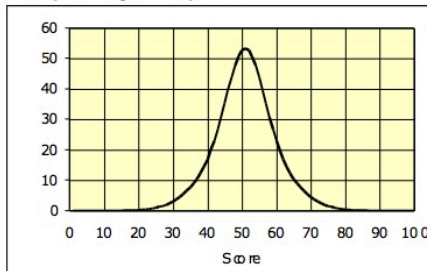
Frequency Graph A



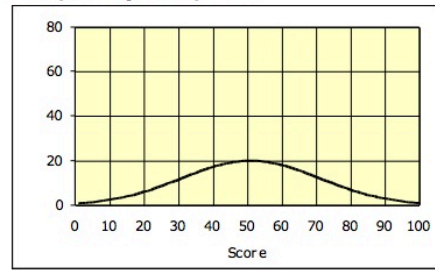
Frequency Graph B



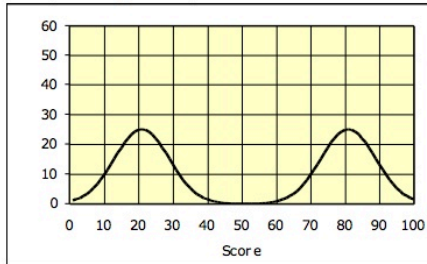
Frequency Graph C



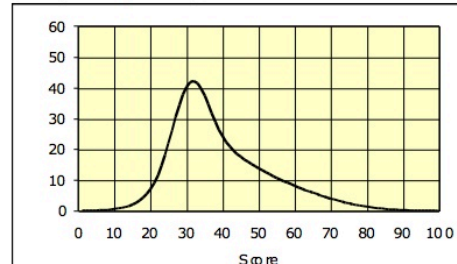
Frequency Graph D



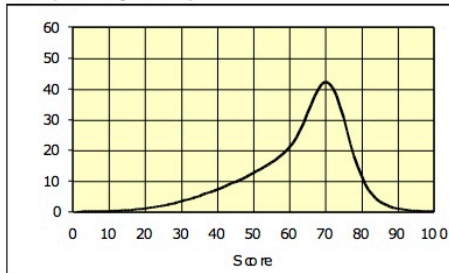
Frequency Graph E



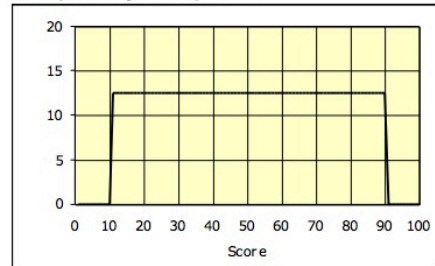
Frequency Graph F



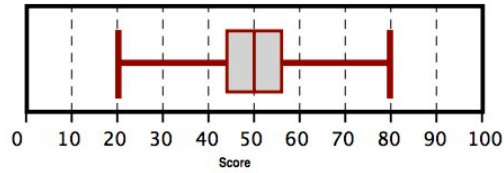
Frequency Graph G



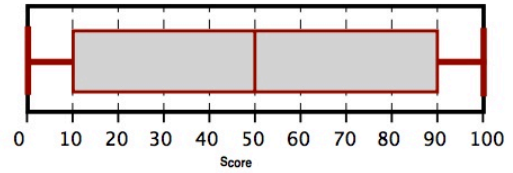
Frequency Graph H



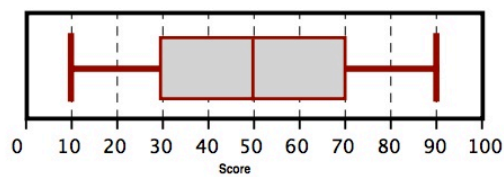
Box Plot 1



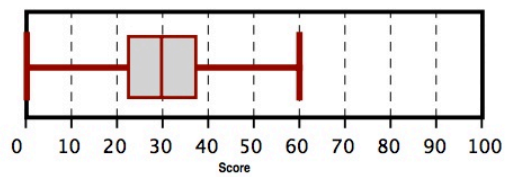
Box Plot 2



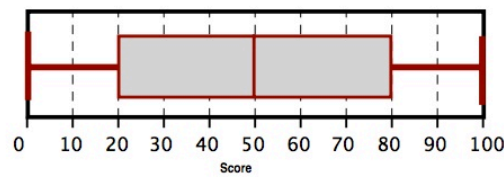
Box Plot 3



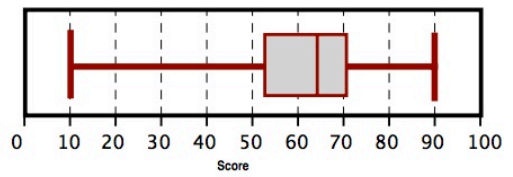
Box Plot 4



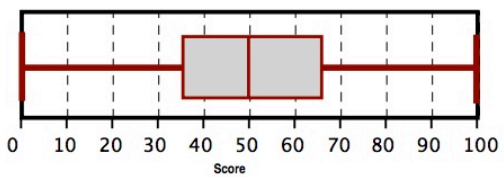
Box Plot 5



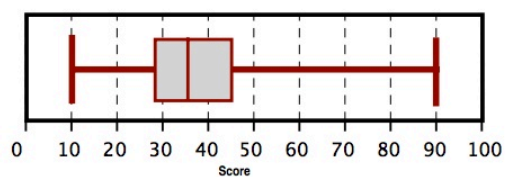
Box Plot 6



Box Plot 7



Box Plot 8



§ 2.3 Frequentieverdelingen typeren

§ 2.3.1 Introductie

Kijkend naar een datarepresentatie valt meestal al snel op hoe de verdeling van de tellingen/frequenties over de verschillende waarden eruitziet. Zitten de meeste tellingen juist bij de kleinste waarden of bij de grootste? Of zitten de meeste op een kluitje in het midden? Of nog weer anders?

In deze paragraaf leer je frequentieverdelingen typeren met behulp van kentallen. Dat doe je al heel vaak onbewust.

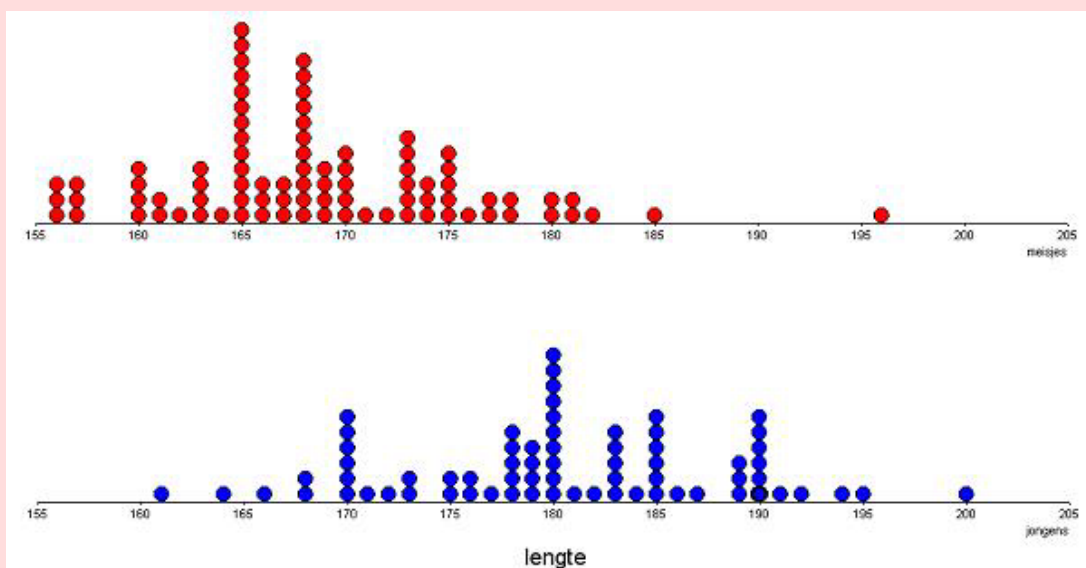
Voorbeelden zijn:

- Deze klas A heeft gemiddeld een 6-, terwijl klas B gemiddeld een 7 heeft.
- Het gemiddelde inkomen in land A is veel lager dan het gemiddelde inkomen in land B.
- Deze speler A is veel constanter in zijn spel dan zijn teamgenoot speler B.

Hierboven blijkt al dat bij de typering van een frequentieverdeling ten minste twee aspecten belangrijk zijn: de ligging (het centrum) en de spreiding.

§ 2.3.2 Centrale vraag

Je ziet hieronder de dotplots van de lengtes van een groep meisjes en van een groep jongens.



Centrale vraag 1

Hoe kun je het centrum van deze twee frequentieverdelingen bepalen?

Centrale vraag 2

Hoe kun je de spreiding van deze twee frequentieverdelingen bepalen?

Centrale vraag 3

Hoe kun je de vorm van deze twee frequentieverdelingen typeren?

Centrale vraag 4

Wat zijn de voor- en nadelen van de verschillende centrum- en spreidingsmaten voor de verschillende typen frequentieverdelingen?

§ 2.3.3 Frequentieverdelingen typeren

Centrummaten van een frequentieverdeling

Opgave 44

Bekijk de dotplots uit de centrale vraag.

- Waar zou je – op het eerste gezicht – bij beide deelgroepen het midden van de frequentieverdeling plaatsen? Licht je antwoord toe.
- Bereken voor beide deelgroepen de gemiddelde lengte.
- De lengte met de grootste frequentie heet de modale lengte. Bepaal de modale lengte van beide deelgroepen.
- De mediaan is de lengte die op de helft van de verdeling zit, dus waar 50 procent van de lengtes onder zit (en dus ook 50 procent erboven). Bepaal de mediaan van de lengtes van de meisjes. Doe dat ook van de jongens.
- Zet voor beide deelgroepen het gemiddelde, de modus en de mediaan op een rijtje van klein naar groot en bespreek de verschillen tussen beide deelgroepen.

Opgave 45

In een dorp wonen tien mensen. Daarvan verdienen negen inwoners maandelijks 1200 euro en één rijke inwoner verdient 20.000 euro per maand.

- Teken een dotplot van de inkomens.
- Schat het gemiddelde op basis van de dotplot.
- Bereken het gemiddelde en geef die met een verticale lijn in je dotplot aan.
- Komt het gemiddelde overeen met de schatting uit opgave b?
- Waarom wordt wel gezegd dat het gemiddelde het evenwichtspunt van een verdeling is?
- Bepaal het modale inkomen.
- Hoe groot is de mediaan van de inkomens?
- Vallen mediaan, modus en gemiddelde enigszins samen?





Antwoord op centrale vraag 1

Er zijn drie centrummaten waarmee je het centrum van de frequentieverdelingen kunt bepalen:

- Gemiddelde, gemiddelde lengte.
- Mediaan.
- Modus, modale lengte.

Spreidingsmaten van een frequentieverdeling

Opgave 46

Bekijk opnieuw de dotplots uit de centrale vraag.

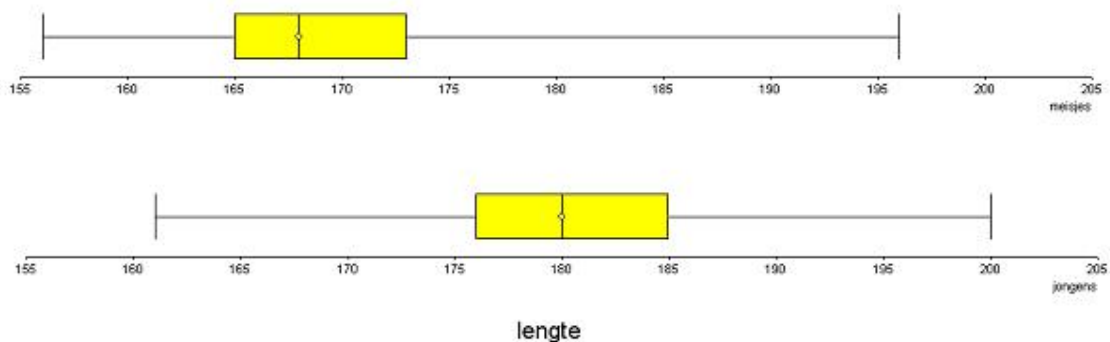
Het verschil tussen de grootste en de kleinste lengte heet de **spreidingsbreedte**.

- Bereken voor de lengte van de jongens de spreidingsbreedte. En voor de meisjes.
- De beide spreidingsbreedten verschillen nauwelijks.
Vind je dat de spreiding van de lengten van de jongens en de meisjes vrijwel even groot is?

Opgave 47

Bekijk weer de dotplots uit de centrale vraag.

- Welke lengtes hebben de 25 procent kleinste jongens?
- Laat met een berekening zien dat 25 procent van de langste meisjes inderdaad de lengten 173 tot en met 197 centimeter hebben.
- Bepaal nu zowel voor de jongens als de meisjes uit de dotplots de grenzen van de vier kwarten in een boxplot. Ga na of deze grenzen de volgende boxplots opleveren.



- Het verschil tussen Q1 en Q3 heet **interkwartielafstand**. Bepaal de interkwartielafstand voor de lengtes van de meisjes en voor de lengtes van de jongens. Wijken de interkwartielafstanden erg van elkaar af?

Antwoord op centrale vraag 2

Er zijn drie spreidingsmaten waarmee je de spreiding van de frequentieverdelingen kunt bepalen:

- Spreidingsbreedte.
- Interkwartielafstand.
- Standaardafwijking.

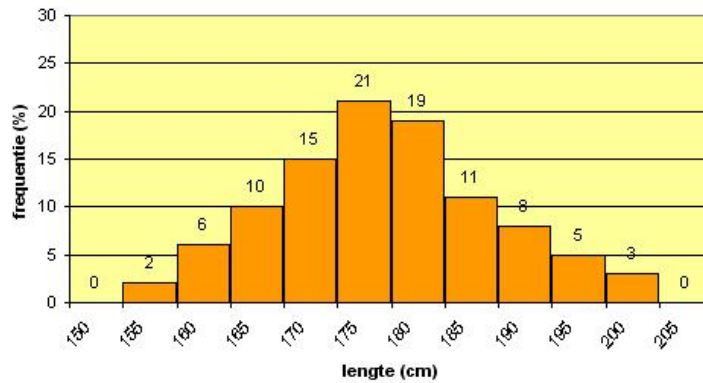
Deze laatste spreidingsmaat komt pas in boekje 3 aan de orde.

Vorm van een frequentieverdeling

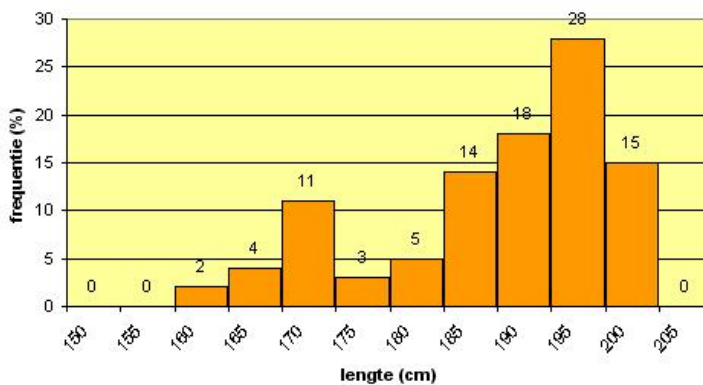
Opgave 48

Hieronder staan staafdiagrammen (histogrammen) van de lengteverdeling van verschillende groepen sporters.

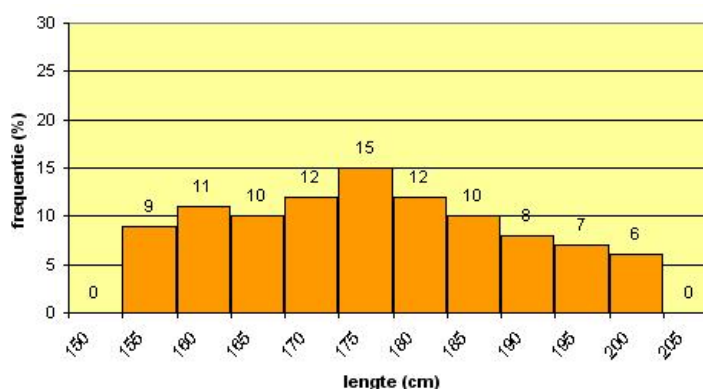
Staafdiagram 1:



Staafdiagram 2:



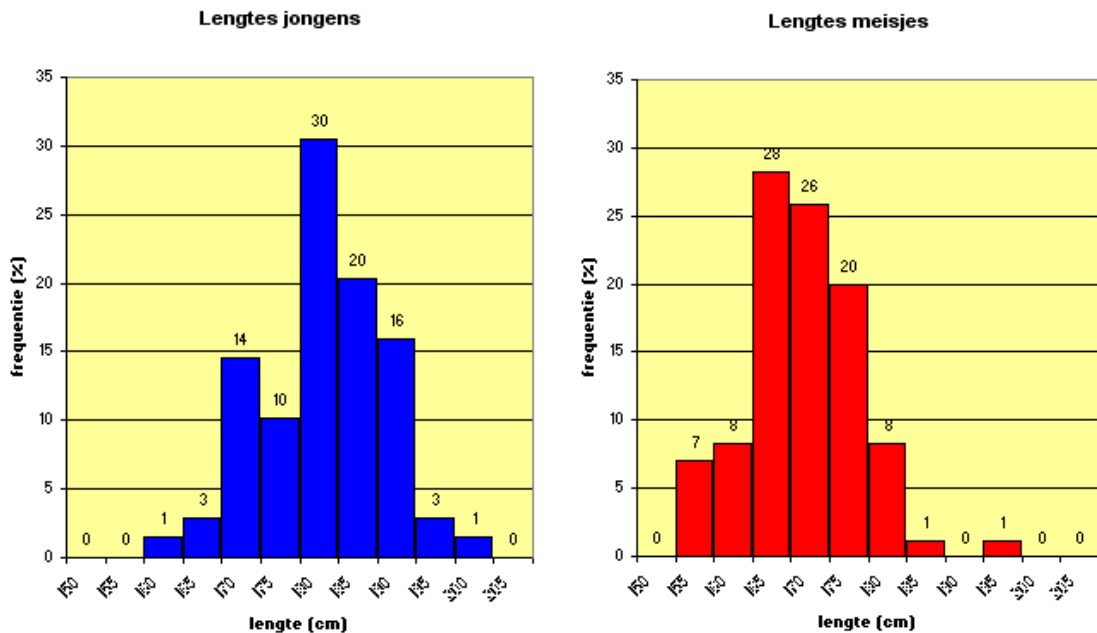
Staafdiagram 3:



- Bij welk staafdiagram is duidelijk sprake van een scheve verdeling? Links-scheef of rechts-scheef?
- Welk staafdiagram is het meest symmetrisch?
- Welk staafdiagram is het meest gelijkmatig en dus uniform?
- Welk staafdiagram is duidelijk tweetoppig?

Opgave 49

Je ziet hier staafdiagrammen (histogrammen) van de lengtes van jongens en meisjes uit de dotplots van de centrale vraag: de lengtes zijn nu ingedeeld in klassen van 5 centimeter breed.



- Geef typeringen voor de frequentieverdeling van de lengtes van de jongens. Kies uit de volgende typeringen: uniform (gelijkmatig), symmetrisch, links-scheef, rechts-scheef, eentoppig, tweetoppig, meertoppig, klokvormig (zoals een kerstklok).
- Doe dit ook voor de frequentieverdeling van de lengtes van de meisjes.
- Welke typeringen zijn hetzelfde voor beide verdelingen? Welke verschillend?

Antwoord op centrale vraag 3

Je kunt de **vorm** van frequentieverdelingen typeren met, onder meer, de volgende termen:

- Uniform (gelijkmatig).
- Symmetrisch (rond het midden).
- Klokvormig (zoals een kerstklok).
- Scheef (links-scheef of rechts-scheef).
- Eentoppig, tweetoppig, meertoppig.

Centrum- en spreidingsmaten versus vorm

Opgave 50

Bekijk opnieuw de staafdiagrammen (histogrammen) van de lengteverdeling van verschillende groepen sporters in opgave 48.

- Bepaal voor de drie staafdiagrammen het gemiddelde, de mediaan en de modus.
Geef voor ieder van de frequentieverdelingen aan welke centrummaat de verdeling het beste typeert en leg uit waarom.
- Bepaal voor de drie staafdiagrammen de spreidingsbreedte en de interkwartielafstand.
Geef voor ieder van de frequentieverdelingen aan welke spreidingsmaat de verdeling het beste typeert en leg uit waarom.

Verwijder in het tweede staafdiagram (dat rechts-scheef maar ook tweetoppig verdeeld is) de tweede, lagere top door de frequentie van de staaf te wijzigen in 3 in plaats van 11. De scheve verdeling heeft nu een *staart* die langzaam afloopt.

- Bepaal opnieuw het gemiddelde, de mediaan en de modus en vergelijk ze met de tweetoppige situatie. Verandert je idee over de beste centrummaat voor deze scheve frequentieverdeling?

Verander nogmaals dezelfde staaf in het tweede staafdiagram: hij had oorspronkelijk frequentie 11, kreeg toen frequentie 3 en nu geef je hem frequentie 30.

- Bepaal wederom het gemiddelde, de mediaan en de modus en vergelijk ze met de vorige twee situaties. Verandert je idee over de beste centrummaat voor deze scheve frequentieverdeling?

Opgave 51

In opgave 45 heb je de centrummaten berekend van de inkomens van een dorp met tien inwoners. De rijke inwoner is ondertussen verhuisd naar de stad.

- Welke centrummaten zullen hierdoor veranderen? Beargumenteer je antwoord.
- Denk je dat de interkwartielafstand zal veranderen? Zo ja: hoe? Zo nee: waarom niet?
- Verandert de spreidingsbreedte hierdoor?
- Kun je je voorstellen dat het inkomen van de rijke man een zogenaamde *uitschieter* binnen de oorspronkelijke inkomensverdeling is? Leg uit.

Antwoord op centrale vraag 4

Het nut van een centrum- of spreidingsmaat is afhankelijk van de vorm van een frequentieverdeling. Het gemiddelde en de spreidingsbreedte bijv. zijn gevoelig voor uitschieters en (lange) staarten in een verdeling: je kunt dan beter mediaan en interkwartielafstand kiezen om je frequentieverdeling zinvol mee te typeren.

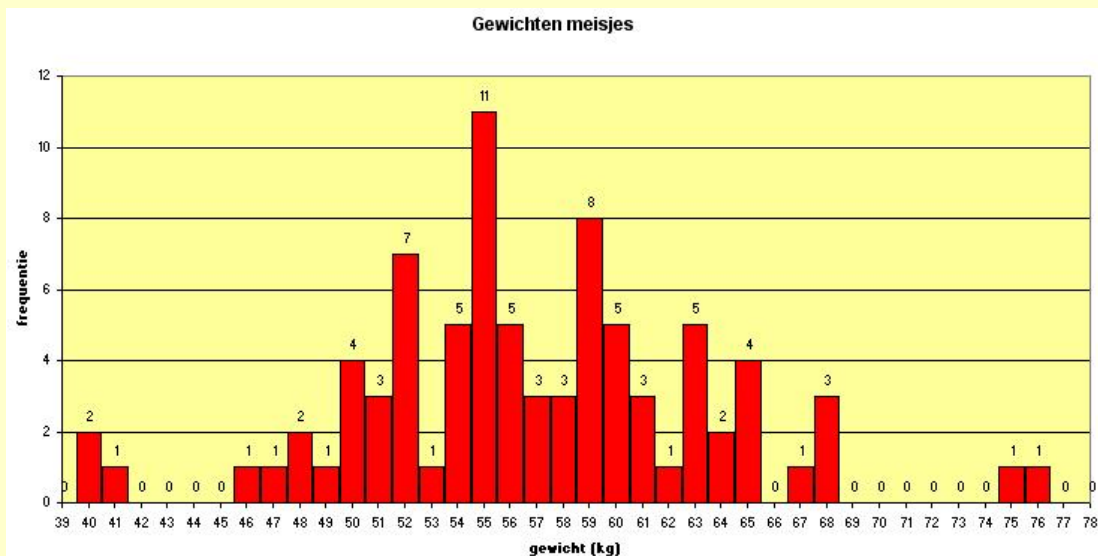
Als de hoogste top (modus) van een meertoppige verdeling bijvoorbeeld in de staart zit, dan is de modus geen zinvolle typering voor deze frequentieverdeling.



§ 2.3.4 Oefenen

Voorbeeld

Je ziet hier een staafdiagram met de gewichten van de meisjes.



Bereken de mediaan en het gemiddelde van de gewichten in 1 decimaal nauwkeurig. Bereken ook de spreidingsbreedte en de kwartielafstand. Ga na welke van deze centrum- en spreidingsmaten het meest zinvol is.

Uitwerking

De mediaan verdeelt de gewichten in twee gelijke delen (ze staan al op volgorde). Omdat er 84 meisjes zijn die hun gewicht hebben opgegeven, neem je hiervoor het gemiddelde van het 42^e en het 43^e gewicht.

Het 42^e gewicht is 56 kilogram en het 43^e ook, dus de mediaan is 56 kilogram.

Het gemiddelde gewicht bereken je met behulp van een frequentietabel.

Je maakt dan een extra kolom met gewicht x frequentie.

Het gemiddelde wordt $\frac{4771}{84} \approx 56,8$ kg.

De spreidingsbreedte is hier $76 - 40 = 36$ kg.

Voor de kwartielafstand moet je beide kwartielen Q1 en Q3 bepalen.

Q1 verdeelt de eerste helft van de gewichten weer in twee gelijke delen en is dus het gemiddelde van het 21^e en het 22^e gewicht.

Dus Q1 = 52 kg. En op dezelfde manier is Q3 = 60 kg.

De kwartielafstand is daarom $60 - 52 = 8$ kg.

Maar goed dat je deze getallen in het vervolg meestal door de computer laat berekenen.

gewicht	freg	
39	0	0
40	2	80
41	1	41
42	0	0
43	0	0
44	0	0
45	0	0
46	1	46
47	1	47
48	2	96
49	1	49
50	4	200
51	3	153
52	7	364
53	1	53
54	5	270
55	11	605
56	5	280
57	3	171
58	3	174
59	8	472
60	5	300
61	3	183
62	1	62
63	5	315
64	2	128
65	4	260
66	0	0
67	1	67
68	3	204
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	1	75
76	1	76
		4771

Hoe zinvol zijn nu al die maten?

De modale lengte zegt niet veel over de verdeling. In dit geval zit die lengte nog redelijk in het midden, maar dat is toeval. Juist de waarden die meer in het midden zitten komen weinig voor.

De mediaan is een zinvolle maat: 50 procent van de lengtes zit er onder en 50 procent zit er boven.

Ook het gemiddelde is hier een zinvolle maat: in dit geval met die gewichten is het letterlijk het evenwichtspunt van de verdeling.

De kwartielafstand is als maat voor de spreiding ook geschikter dan de spreidingsbreedte: die laatste maat wordt nogal bepaald door de uitschieters bij deze verdeling. Dat geldt niet voor de kwartielafstand.

Opgave 52

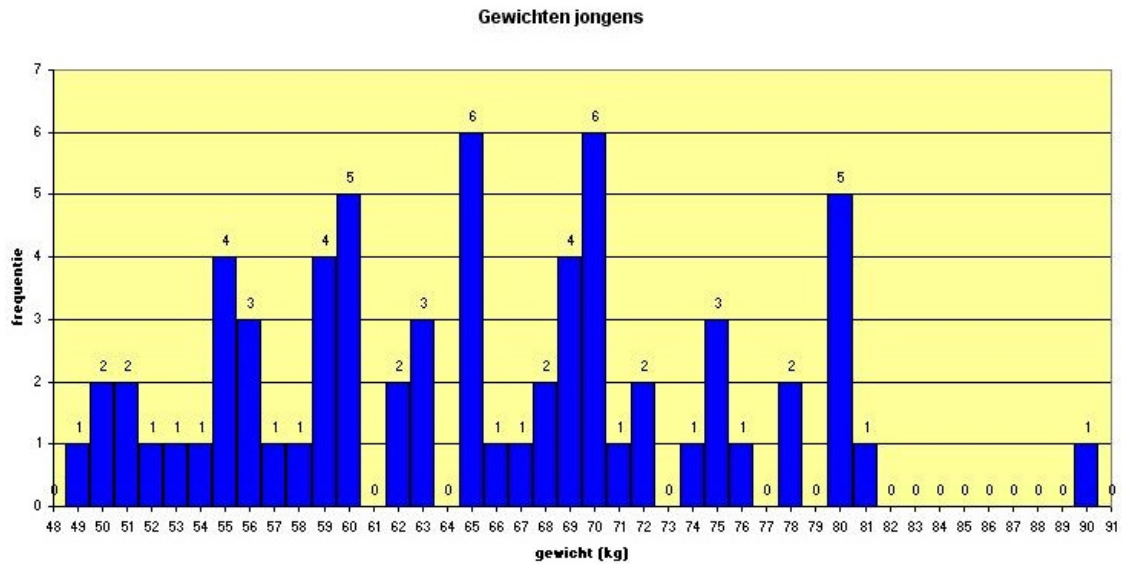
Bekijk de dotplots van de lengtes van de jongens en de meisjes uit de centrale vraag nog eens. In opgave 44 en 46 heb je van beide deelgroepen de centrum- en spreidingsmaten bepaald.

- Welke centrummaat en welke spreidingsmaat geven de datasets van de deelgroepen het beste weer?
- Laat zien dat bij de jongens de waarden 161 en 200 centimeter uitschieters zijn.
- Laat deze data weg en maak een nieuw overzicht van de drie centrummaten en de twee spreidingsmaten.
- Welke spreidingsmaat wordt door deze uitschieters sterk beïnvloed en welke niet?
- Welke centrummaat wordt door deze uitschieters sterk beïnvloed?
- Vind je het verantwoord om uitschieters weg te laten bij het samenvatten van een frequentieverdeling? Geef argumenten voor en tegen.



Opgave 53

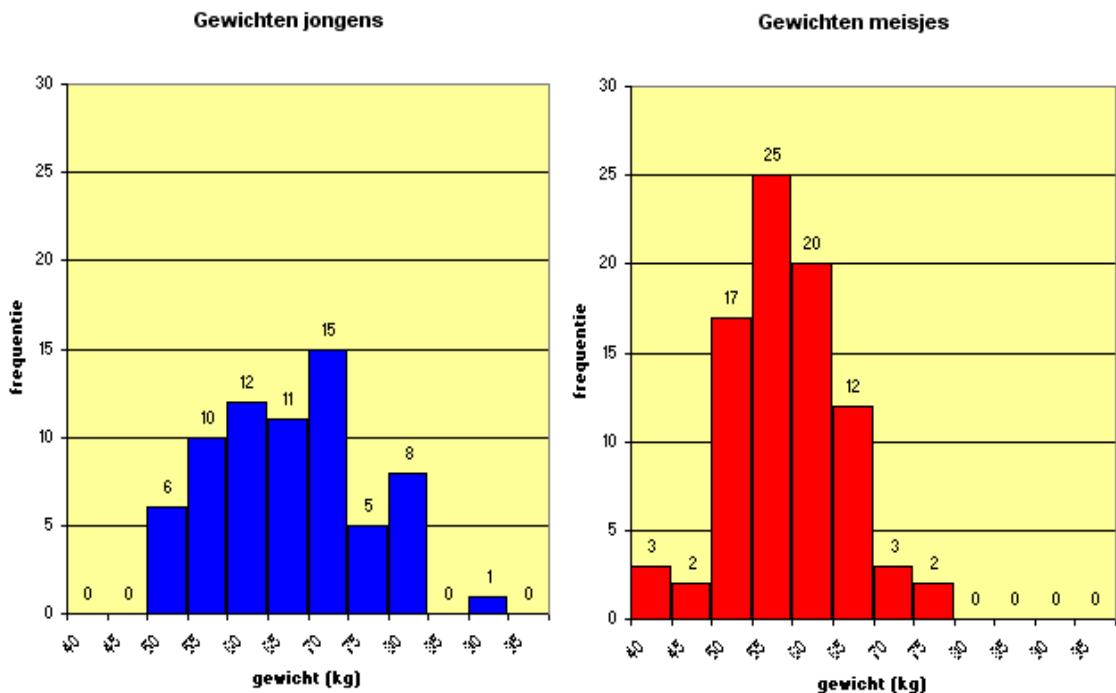
Bekijk het staafdiagram voor de gewichten van de jongens.



- Bereken de mediaan en het gemiddelde van de gewichten van de jongens.
- Waarom is nu de modus niet eens vast te stellen?
- Bepaal de spreidingsbreedte en de kwartielafstand.
- Er is bij de jongens één uitschieter.
Welke centrummaat en/of spreidingsmaat verandert het sterkst als je deze uitschieter weglaat?
- Veranderen de centrum- en/of de spreidingsmaten als je alle absolute frequenties omrekent naar relatieve frequenties?
- Tussen welke waarden wegen de 25 procent lichtste jongens?
- Hoeveel procent van de jongens weegt meer dan 78 kilogram?



Opgave 54

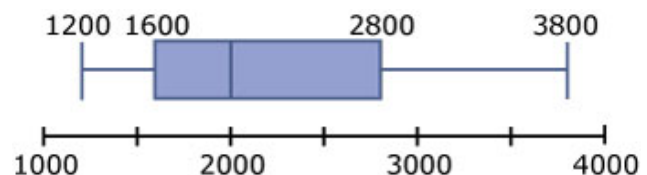


Je ziet hier de frequentieverdelingen van de gewichten van de jongens en de meisjes. Ze zijn elk gegroepeerd in klassen. De vraag is of je de centrummaten dan nog kunt berekenen.

- Waarom kun je vanuit deze frequentieverdelingen de mediaan niet meer vaststellen?
In welke klasse zit de mediaan bij de meisjes? En bij de jongens?
- Maak bij deze klassenindeling frequentietabellen voor de gewichten van de jongens en de meisjes en voeg daaraan de klassenmiddens toe.
- Waarom kun je met deze klassen het gemiddelde alleen nog maar schatten?
- Geef een schatting van het gemiddelde met behulp van de klassenmiddens zowel voor de jongens als voor de meisjes.
- Denk je dat je antwoorden afwijken van de daadwerkelijke gemiddelden?

Opgave 55

In een bedrijf met 120 medewerkers is het modale salaris ongeveer 1600 euro per maand. Het gemiddelde salaris is 1800 euro per maand. Het hoogste salaris is dat van de algemeen directeur. De boxplot vat de verdeling van de salarissen samen.

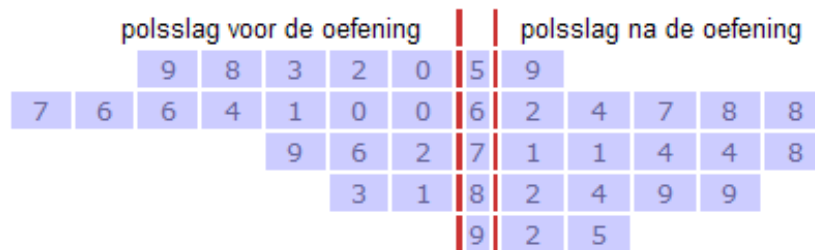


Bereken in de volgende gevallen telkens weer het modale salaris en het gemiddelde salaris en teken de nieuwe boxplot.

- Alle medewerkers krijgen een loonsverhoging van 3 procent.
- Alle medewerkers krijgen een maandelijkse toeslag van 200 euro.
- Het salaris van de algemeen directeur wordt met 800 euro per maand verhoogd.

Opgave 56

Als je in de sportzaal een tijdje een bepaalde oefening hebt gedaan, gaat je polsslag omhoog. In dit tweezijdige steelbladdiagram vind je wat data. Van elke sporter werd één keer voor en één keer na de oefening de polsslag gemeten.



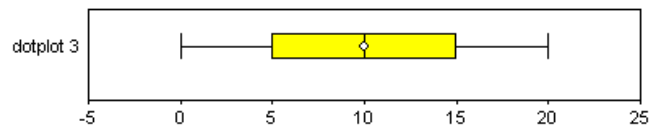
- Waarom zegt de modale polsslag hier weinig over het centrum van de verdeling?
Is de modale polsslag een zinvol getal?
- Bereken de gemiddelde polsslag voor en ook na de oefening.
Is het gemiddelde een bruikbare centrummaat voor deze situatie?
- Bepaal de mediaan en de kwartielen.
Zijn twee boxplots een geschikt middel om beide datasets voor deze situatie te vergelijken?
- Is het wel handig om de polsslag voor en na de oefening apart in beeld te brengen?



Opgave 57

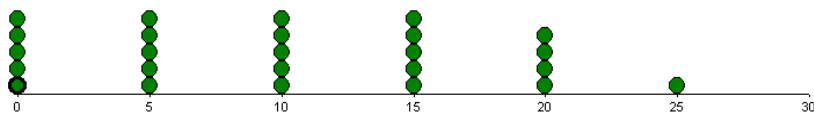
Hieronder zie je een boxplot en een aantal dotplots, die erg van vorm verschillen.

- Verander bij elke dotplot één waarneming van plaats zodat de boxplot de verdeling goed weergeeft.
- Verzin zelf een dotplot die goed weergegeven wordt door deze boxplot.



- Benoem de vorm van de frequentieverdeling in deze boxplot.

Dotplot 1:



Dotplot 2:

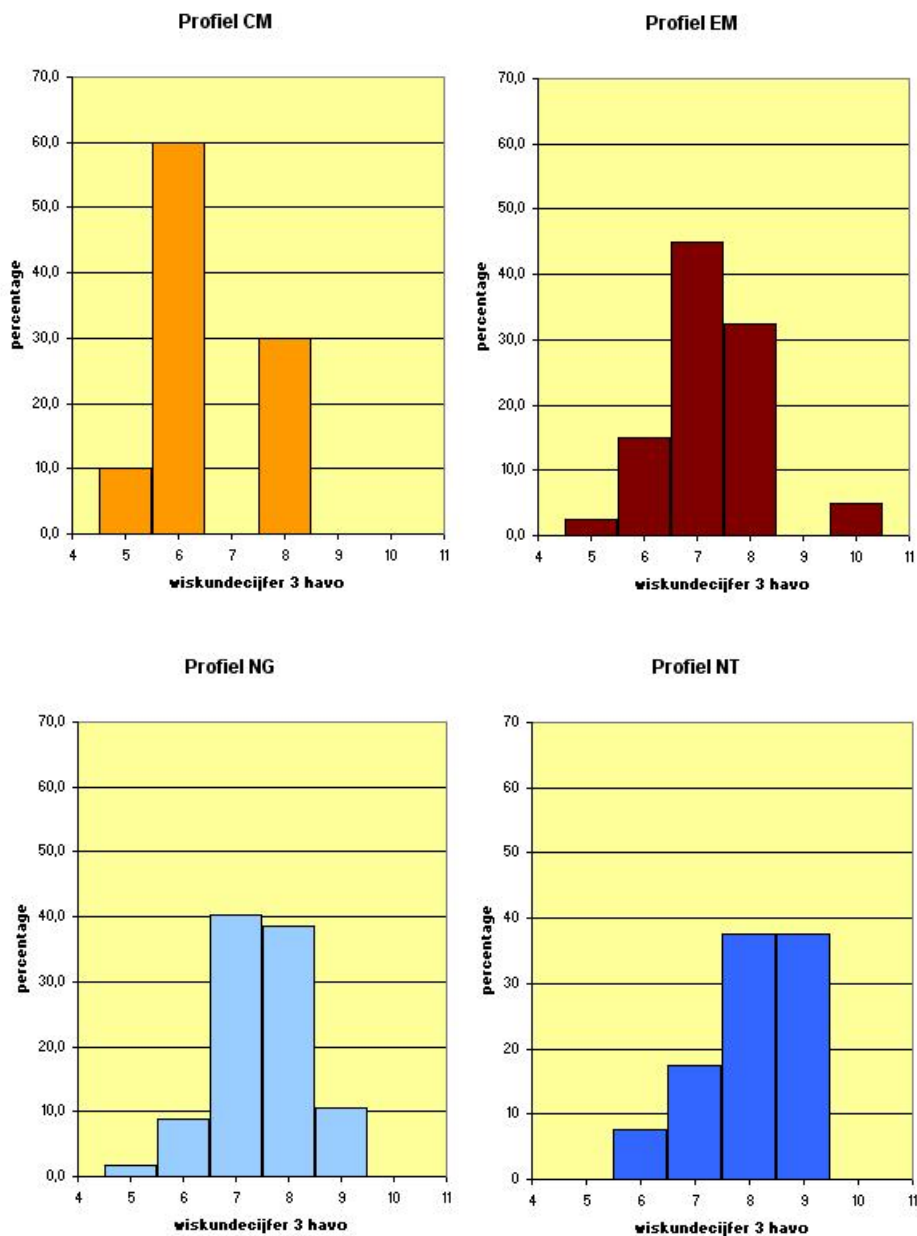


Dotplot 3:



Opgave 58

Je ziet hier staafdiagrammen (histogrammen) van de wiskundecijfers in havo 3 voor de leerlingen in de verschillende profielen.



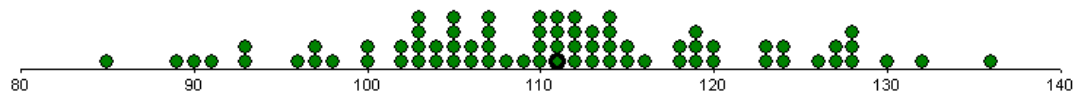
- Beschrijf de vorm van ieder van deze frequentieverdelingen.
- Vergelijk de frequentieverdelingen met elkaar. Probeer conclusies te trekken.
- Bepaal van de vier verdelingen de mediaan en het gemiddelde.
- Bij welke van deze vier verdelingen liggen mediaan en gemiddelde het dichtst bij elkaar?
Kun je dit ook zien aan het staafdiagram?

Opgave 59

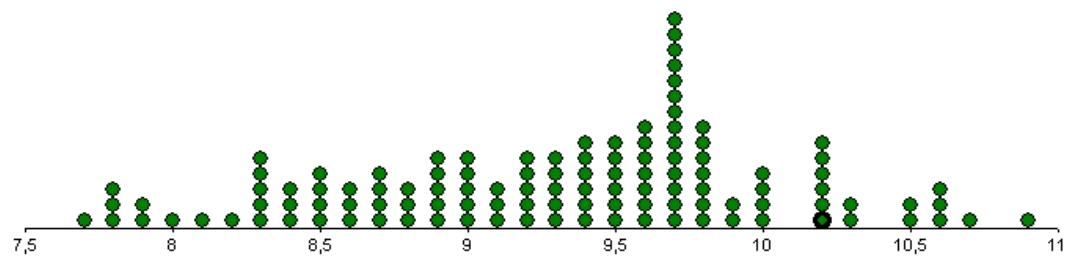
Hieronder en op de volgende pagina zie je dotplots van een aantal datasets.

Beschrijf van elke dataset de vorm van de verdeling.

Ga daarbij elk van de genoemde vormkenmerken na.

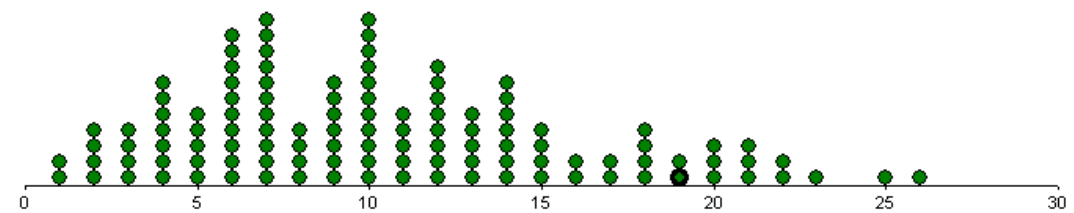


IQ



gemtemp

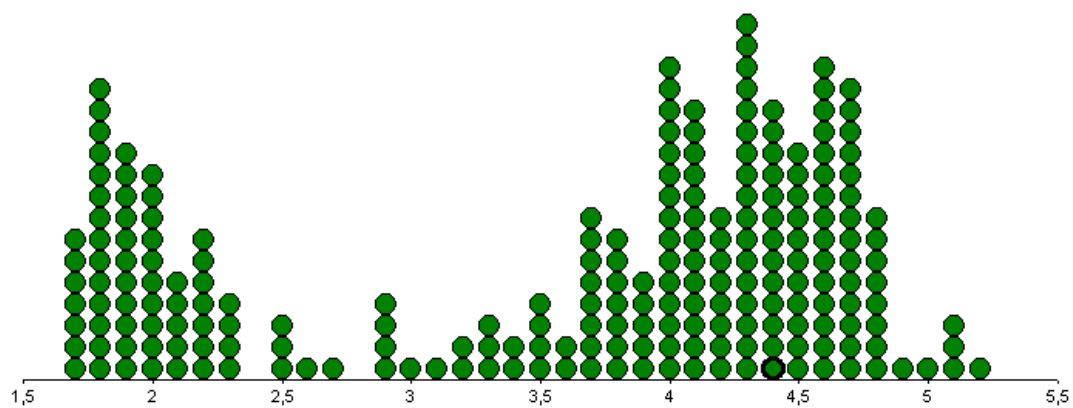
Gemiddelde temperatuur (in °C)



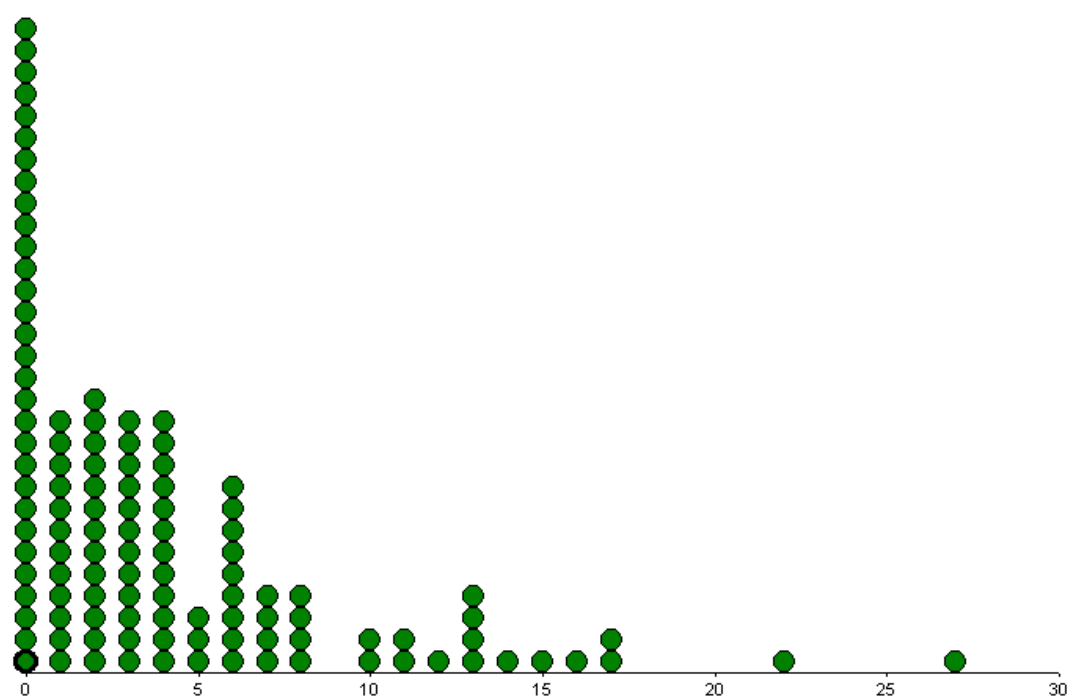
dhagel

Dagen met hagel





Eruptieduur



d-10

Dagen met strenge vorst (≤ -10 °C)



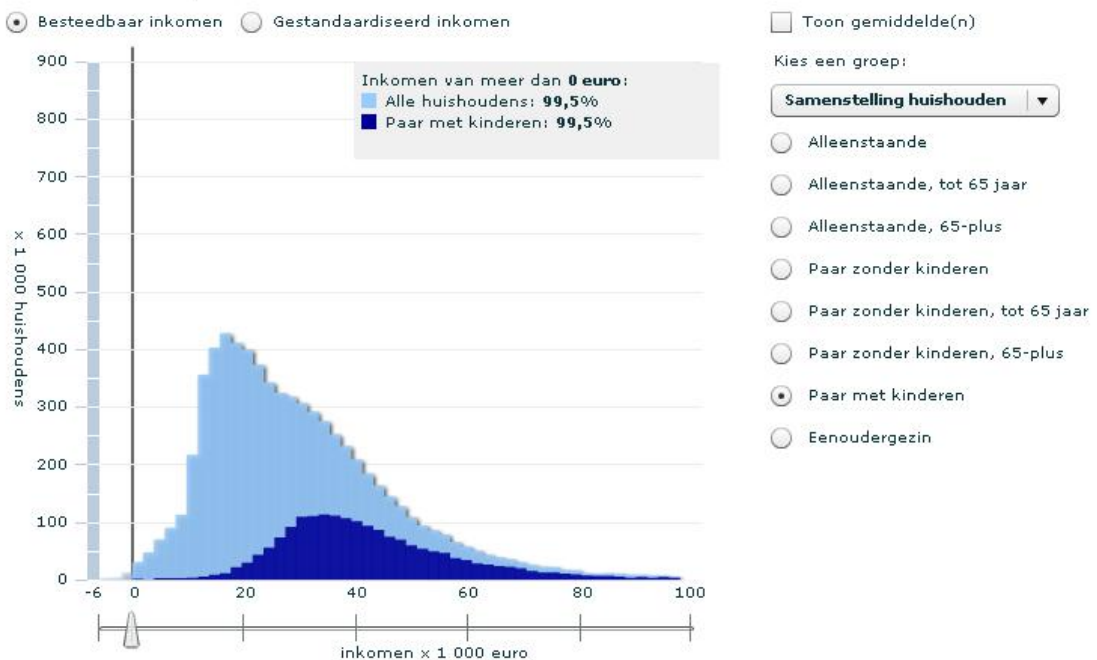
Opgave 60

Je ziet hieronder en op de volgende pagina's drie afbeeldingen van een animatie die is te vinden op de website van het CBS.

Het zijn staafdiagrammen die de inkomensverdeling in Nederland in 2007 in kaart brengen.

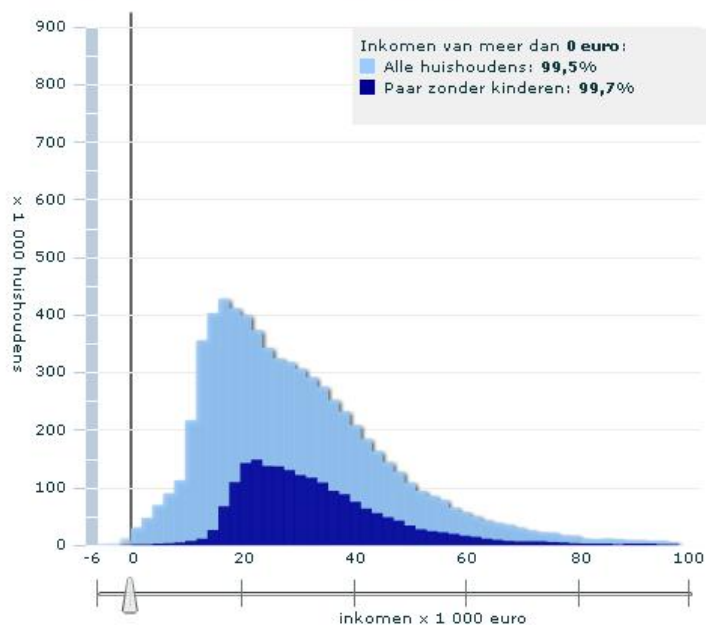
- Beschrijf de vormen van al deze verdelingen.
- Bepaal voor elke verdeling de spreidingsbreedte.
- Ga voor elke verdeling na hoe het modale inkomen, de mediaan en het gemiddelde ten opzichte van elkaar liggen.
- Vergelijk de inkomensverdelingen van de deelgroepen *Paar met kinderen*, *Paar zonder kinderen* en *Alleenstaanden*. Probeer ook een verklaring te geven voor de verschillen.

Inkomensverdeling huishoudens 2007



Inkomensverdeling huishoudens 2007

☒ Besteedbaar inkomen ☐ Gestandaardiseerd inkomen



☐ Toon gemiddelde(n)

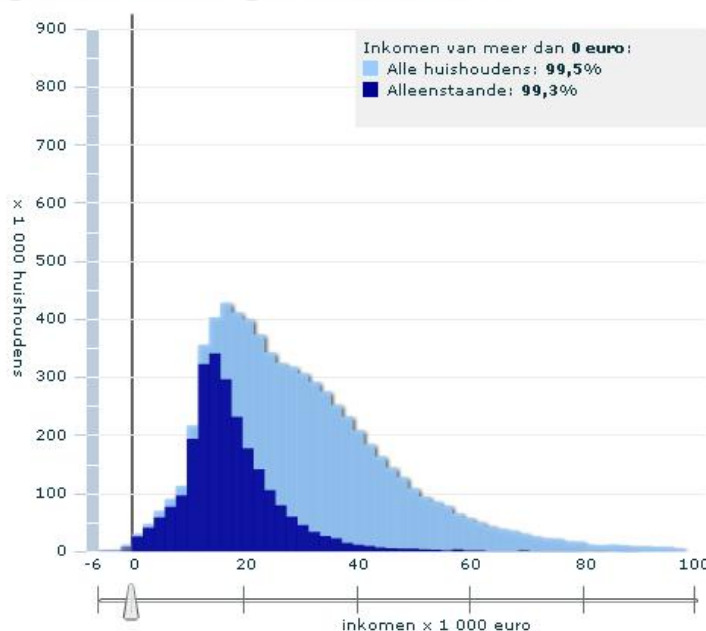
Kies een groep:

Samenstelling huishouden ▼

- ☐ Alleenstaande
- ☐ Alleenstaande, tot 65 jaar
- ☐ Alleenstaande, 65-plus
- ☒ Paar zonder kinderen
- ☐ Paar zonder kinderen, tot 65 jaar
- ☐ Paar zonder kinderen, 65-plus
- ☐ Paar met kinderen
- ☐ Eenoudergezin

Inkomensverdeling huishoudens 2007

☒ Besteedbaar inkomen ☐ Gestandaardiseerd inkomen



☐ Toon gemiddelde(n)

Kies een groep:

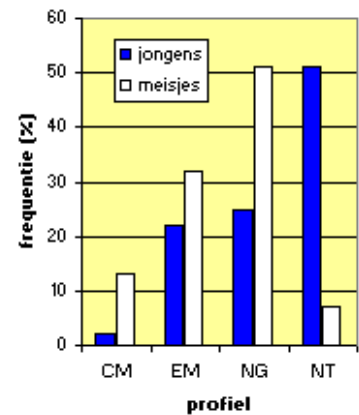
Samenstelling huishouden ▼

- ☒ Alleenstaande
- ☐ Alleenstaande, tot 65 jaar
- ☐ Alleenstaande, 65-plus
- ☐ Paar zonder kinderen
- ☐ Paar zonder kinderen, tot 65 jaar
- ☐ Paar zonder kinderen, 65-plus
- ☐ Paar met kinderen
- ☐ Eenoudergezin

Opgave 61

Dit is een staafdiagram van de profielkeuzes van de groep van 154 leerlingen in havo 4.

- Waarom kun je nu geen spreidingsmaten vaststellen?
- Je kunt wel vaststellen welk profiel de meeste jongens of de meeste meisjes hebben.
Waarom kun je dat toch geen centrummaat noemen?
- Vergelijk nu de profielkeuzes van de meisjes en de jongens.
Wat valt je op?
- Het staafdiagram van de profielkeuzes van de jongens is veel schever dan dat van de meisjes.*
Waarom kun je hier zo'n uitspraak niet doen?



§ 2.3.5 Om te onthouden

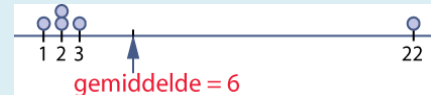
Een frequentieverdeling kun je karakteriseren door:

Centrummaten: getallen die het centrum van de verdeling aangeven.

Spreidingsmaten: getallen die de spreiding van de verdeling weergeven.

Centrummaten zijn:

- **Gemiddelde:** het evenwichtspunt van de verdeling.
- **Mediaan:** de middelste waarde van de verdeling.
- **Modus:** de meest voorkomende waarde.

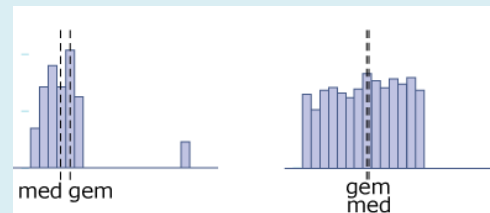


Spreidingsmaten zijn:

- **Interkwartielafstand:** $Q3 - Q1$, (Engels: *Inter Quartile Range IQR*).
- **Spreidingsbreedte:** maximum-minimum.

Bij de **vorm** van een frequentieverdeling let je op:

- Symmetrie.
- Scheefheid.
- Aantal toppen.
- Staart.
- Uitschieters.
- Gelijkmatigheid.



Bij een symmetrische verdeling vallen mediaan en gemiddelde vrijwel samen.



Een **uitschieter** is een waarde die meer dan 1,5 keer de kwartielafstand onder het eerste kwartiel of boven het derde kwartiel zit.

Het nut van een centrum- of spreidingsmaat is afhankelijk van de vorm van een frequentieverdeling.

§ 2.3.6 Geïntegreerd oefenen

📄 Opgave 62

Voor een bepaalde toets kun je maximaal 100 punten scoren.

Hieronder zie je hoe een groep van 40 personen de toets heeft gemaakt.

59 57 53 60 63 58 77 33 50 59 58 75 62 54 53 78 59 68 65 62
57 60 80 47 90 30 60 35 57 87 63 65 63 58 65 70 73 58 63 55

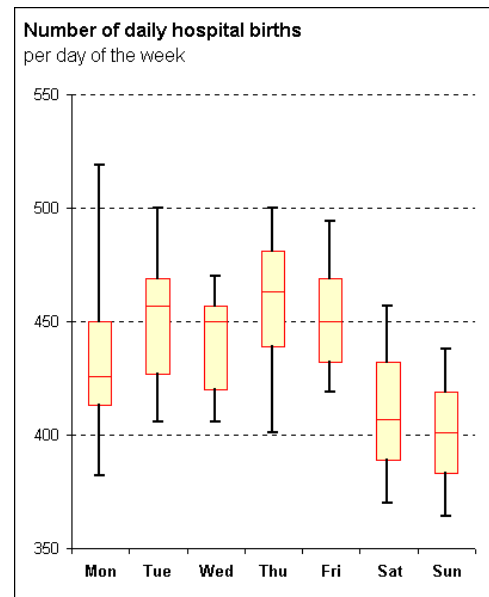
- Hoeveel bedraagt de gemiddelde score in 1 decimaal nauwkeurig?
- Teken de boxplot bij deze scores.
- Welke centrummaat vat de data het beste samen?

- d. Leg uit dat de schatting van het gemiddelde steeds onnauwkeuriger wordt als je de klassenbreedte vergroot.

Opgave 63

Je ziet hier boxplots van het aantal geboorten in ziekenhuizen per dag voor de verschillende dagen van de week.

- Op welke dag van de week is de spreidingsbreedte van het aantal geboortes in ziekenhuizen het grootst? Waarom kun je de dagen niet goed vergelijken met behulp van de spreidingsbreedten?
- Op welke dag van de week is de kwartielafstand van het aantal geboortes in ziekenhuizen het grootst?
- Hoeveel procent van de zondagen zijn er minder dan 400 geboortes in ziekenhuizen?
- Vergelijk de maandag en de vrijdag. Van beide dagen zijn er 52 per jaar. Op welk van deze dagen zijn er in één jaar de meeste bevallingen? Licht je antwoord toe.
- Leg uit waarom het mogelijk is dat het modale aantal bevallingen per dag voor elk van deze dagen hetzelfde is.
- Is het ook mogelijk dat het gemiddelde aantal bevallingen per dag voor elk van deze dagen gelijk is? Licht je antwoord toe.



Opgave 64

Gebruik het bestand ► **SPORTPRESTATIES**. Je vindt er gegevens van brugklassers op sportgebied.

- Bereken voor het vergooien alle centrummaten en alle spreidingsmaten vanuit de ruwe data.
- Waarom kun je dit altijd beter vanuit de ruwe data doen dan vanuit een klassenindeling?
- Probeer conclusies te trekken over het vergooien.

Gebruik daarbij de centrum- en de spreidingsmaten en de vorm van de verdeling. Vermeld ook vooral welke centrum- en welke spreidingsmaten hier zinvol zijn.



§ 2.4 Twee groepen vergelijken

§ 2.4.1 Introductie

Zijn jongens langer dan meisjes?

Hebben leerlingen met een NT-profiel in de derde klas een hoger cijfer voor wiskunde dan leerlingen met een CM-profiel?

Is de huizenprijs in Amsterdam hoger dan in Noordoost-Groningen?

Is de spreiding van de temperatuur in de tropen kleiner dan de spreiding in temperatuur bij ons?

Bij dit soort vragen wil je twee groepen vergelijken, bijvoorbeeld jongens-meisjes, CM-NT-profiel, Amsterdam-Noordoost-Groningen, tropen-Nederland.

In deze paragraaf gaan we dieper in op de vraag hoe je twee groepen kunt vergelijken en waar je dan naar kunt kijken.

§ 2.4.2 Centrale vraag

Centrale vraag 1

We kijken naar de gegevens van 84 leerlingen aan het einde van atheneum 5. Wat kun je zeggen over het verschil in het eindcijfer voor Engels tussen de CM-groep en de NT-groep?

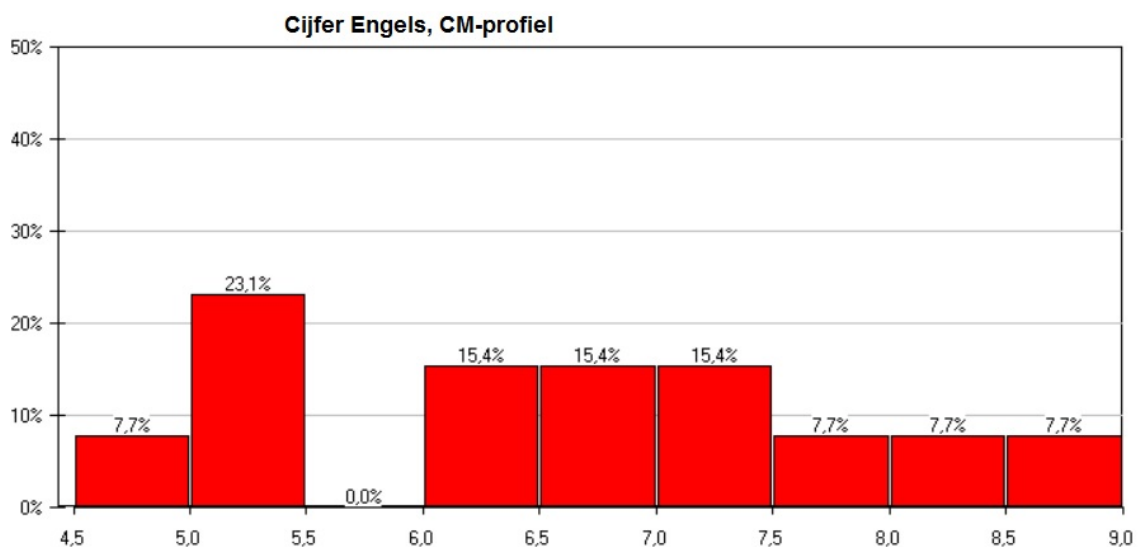
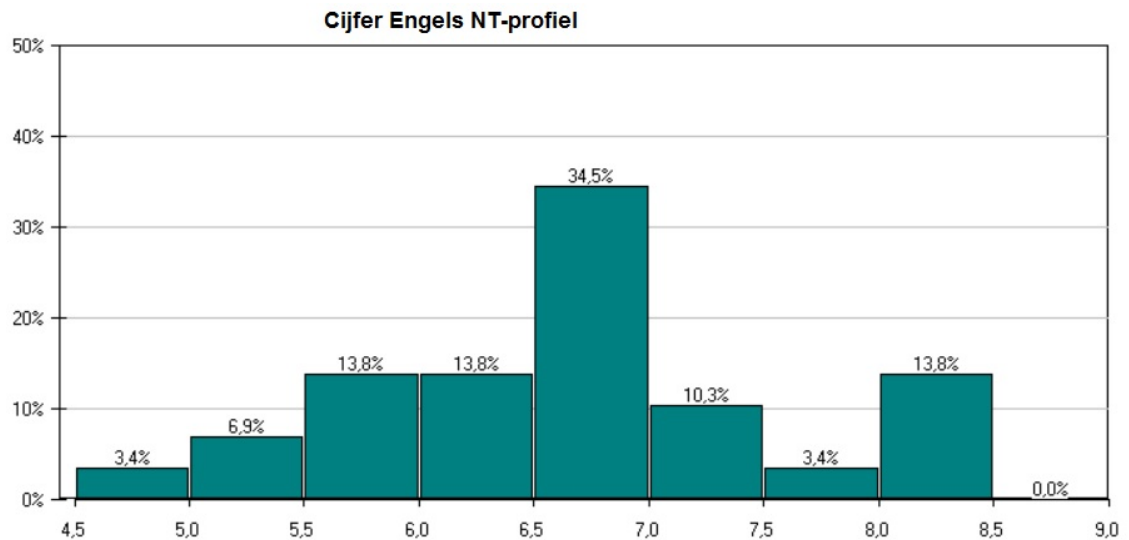
Centrale vraag 2

Welke datarepresentaties kun je gebruiken om de verschillen tussen twee groepen inzichtelijk te maken?



§ 2.4.3 Twee groepen vergelijken

De volgende staafdiagrammen komen uit de dataset ►5A_{TH}. Je zou je kunnen afvragen of leerlingen met een CM-profiel beter zijn in Engels dan de leerlingen met een NT-profiel.



Opgave 65

Je ziet hierboven de staafdiagrammen van de cijfers voor Engels van de twee genoemde profielen.

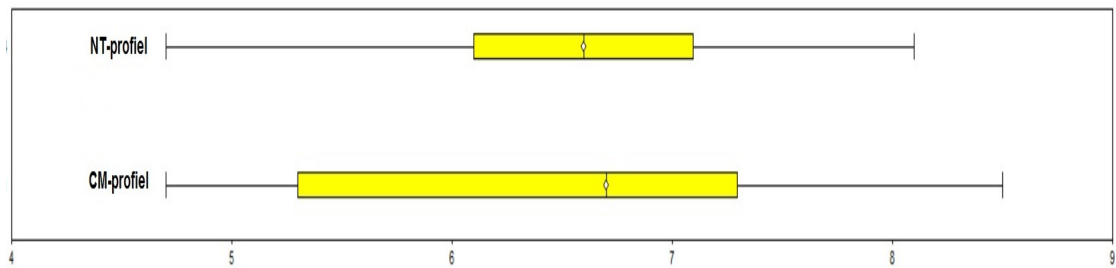
- Beschrijf de verschillen in vorm van deze frequentieverdelingen op basis van het geleerde in paragraaf 3. Wat valt je daarbij op?
- Vergelijk deze frequentieverdelingen. Probeer conclusies te trekken.

Opgave 66

- Bepaal met behulp van ICT van beide verdelingen de mediaan en het gemiddelde.
- Trek op basis van de gevonden waarden van mediaan en gemiddelde een conclusie.

Opgave 67

Hieronder zie je de boxplots die horen bij de beide verdelingen.



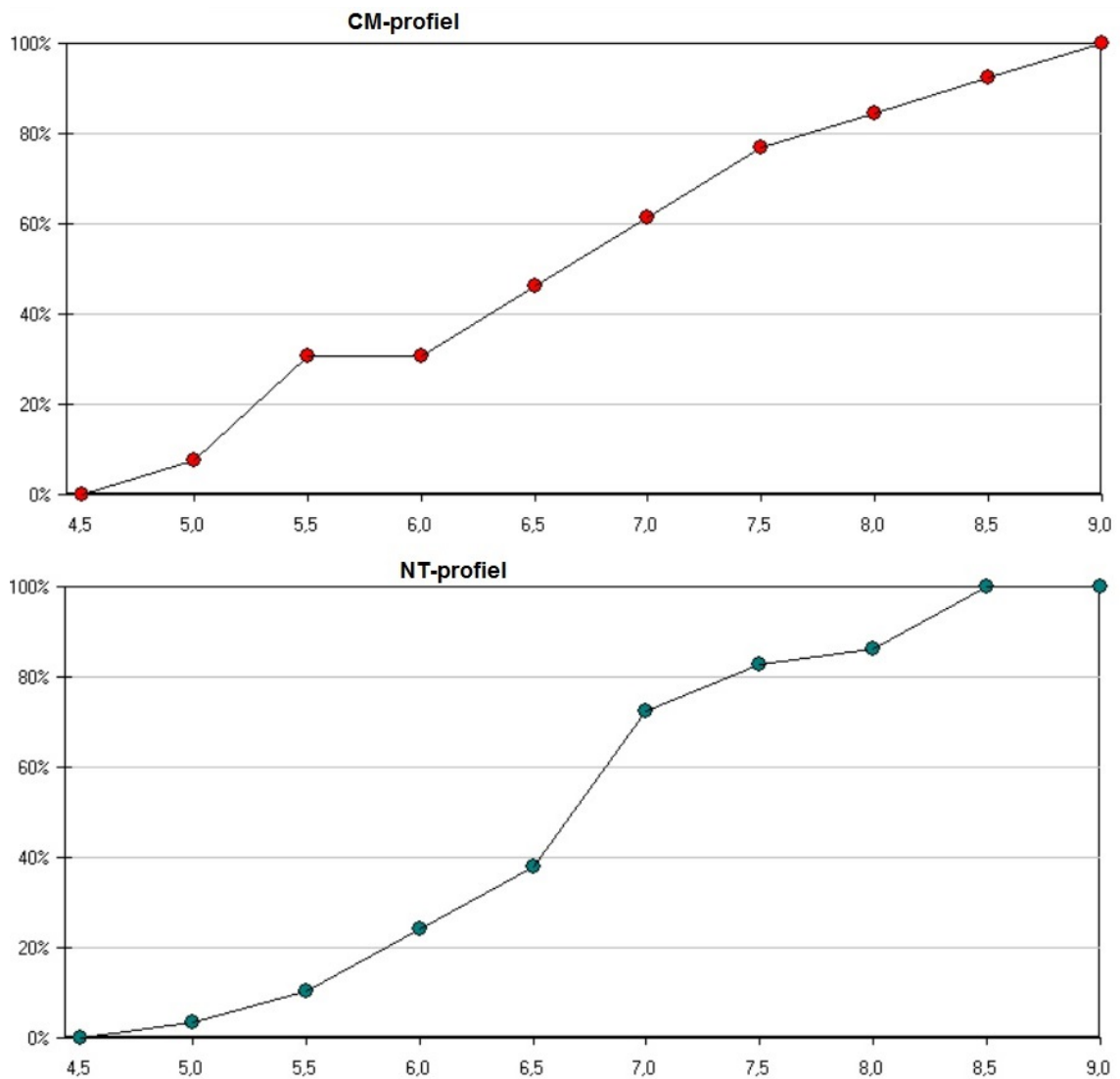
- Wat valt op aan de beide boxplots?
- Trek een conclusie over de verschillen in spreiding van het cijfer voor Engels tussen de beide profielen.



Opgave 68

Hieronder zie je de relatieve cumulatieve frequentiepolygonen die horen bij de beide verdelingen.

- Wat valt je op als je kijkt naar beide grafieken?
- Welke conclusie kun je trekken op basis van deze grafieken?
- Combineer je bevindingen uit de voorgaande opgaven en beschrijf de verschillen in de cijfers voor Engels tussen de leerlingen met een CM-profiel en een NT-profiel.



Antwoord op centrale vraag 1

Het percentage onvoldoendes voor Engels is bij de leerlingen in het CM-profiel hoger dan bij de leerlingen in het NT-profiel (resp. 30,8 en 10,3 procent).

De verdeling over de cijfers is bij de leerlingen in het CM-profiel gelijkmatiger dan bij de leerlingen in het NT-profiel. Bij de laatste groep heeft een groot deel van de leerlingen een cijfer tussen de 6,5 en 7. Bij de leerlingen in het CM-profiel zijn de verschillen in frequentie tussen de cijfers veel kleiner. Dit zie je terug in de boxplots, de interkwartielafstand bij de leerlingen in het CM-profiel is groter dan de kwartielafstand bij leerlingen in het NT-profiel.

Echt hoge cijfers (boven de 8,5) komen alleen voor bij leerlingen in het CM-profiel.

Kanttekening bij deze conclusies

We hebben hier alleen gekeken naar de variabele *cijfer voor Engels*. Eigenlijk zou je ook moeten kijken naar het algemene beeld van de leerlingen. Doen ze het op alle vakken ongeveer even goed, zijn er leerlingen die duidelijk uitvallen? Naar dat soort dingen kijken we in de volgende paragraaf.

Antwoord op centrale vraag 2

Zoals je in het voorgaande hebt gezien, kun je van veel verschillende representaties gebruik maken om verschillen tussen groepen inzichtelijk te maken. Iedere representatie voegt iets toe, het is nooit voldoende om slechts naar één representatie te kijken.

Staafdiagrammen of dotplots geven een goede eerste indruk. Daarna kun je kijken naar boxplots, cumulatieve frequentiepolygonen en de kentallen voor centrum en spreiding.

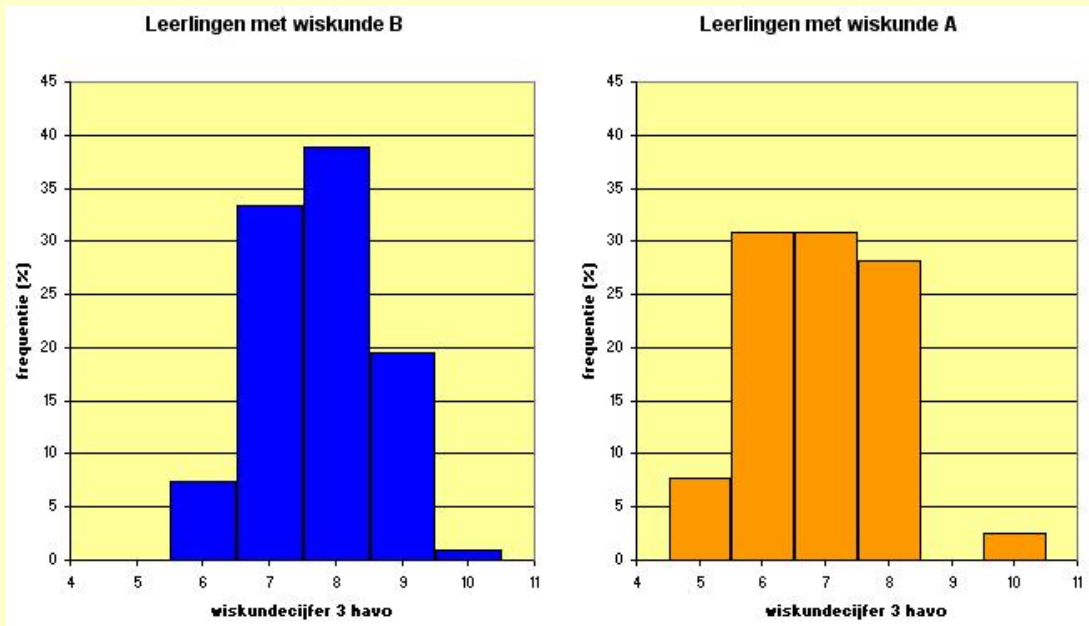


§ 2.4.4 Oefenen

Voorbeeld

Bekijk de dataset ► **GEGEVENS154LEERLINGEN** nog een keer.

Je wilt nagaan of leerlingen die wiskunde B kiezen beter waren in wiskunde in de onderbouw dan leerlingen die wiskunde A kiezen. Daartoe bekijk je de variabele *cijfwis* (het eindcijfer voor wiskunde in havo 3) voor elk van deze deelgroepen.

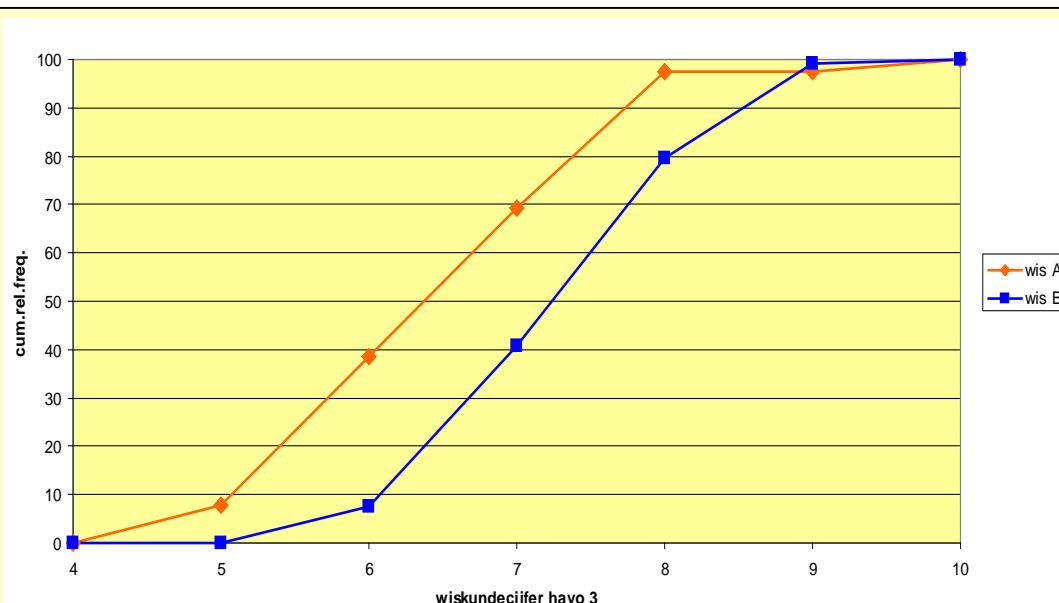


Het linker diagram lijkt redelijk symmetrisch met als top het cijfer 8.

De mediaan van deze gegevens is 8 en het gemiddelde is 7,7.

Het rechterdiagram is in het midden meer gelijkmatig en er is geen echte top. De mediaan van deze gegevens is 7 en het gemiddelde 6,9. Het cijfer 10 wijkt behoorlijk veel af van de andere cijfers, maar is nog net geen uitschieter.

Op de volgende pagina staan de somfrequentiepolygonen van beide verdelingen.



Nu zie je dat de wiskunde B-leerlingen stelselmatig hogere cijfers hebben (op de uitschieter na). Ongeveer 40 procent van de A-leerlingen had bijvoorbeeld een wiskundecijfer van 6 of lager, tegen nog geen 10 procent van de B-leerlingen.

Opgave 69

Bekijk het voorbeeld. Alle percentages zijn gehele getallen.

- Waarom kun je geen van beide verdelingen scheef noemen?
- Reken de gemiddelden en de medianen van beide verdelingen na.
- De somfrequenties zijn uitgezet tegen de gehele cijfers 5, 6, 7, ...
Is dat hier correct?
- Ongeveer 70 procent van de leerlingen met wiskunde A heeft een cijfer van 7 of lager.
Hoeveel procent van de leerlingen met wiskunde B heeft zo'n cijfer?



Opgave 70

Je ziet in de tabel hieronder de frequenties (in procenten) van de lengtes van de meisjes en de jongens uit de dataset ► **GEGEVENS154LEERLINGEN**.

Lengte- klassen	Rel. freq. jongens	Rel. freq. meisjes
155-<160	0	7
160-<165	1	8
165-<170	3	28
170-<175	14	26
175-<180	10	20
180-<185	30	8
185-<190	20	1
190-<195	16	0
195-<200	3	1
200-<205	1	0

- Maak zelf cumulatieve relatieve frequentiepolygonen bij de lengtes van de meisjes en de jongens bij deze klassenindeling.
Denk erom dat nu de somfrequenties bij de rechter klassengrenzen horen!
- Vanuit deze somfrequentiepolygonen kun je de medianen en de kwartielen schatten.
Bij welk percentage vind je de mediaan? En de kwartielen?

Opgave 71

In de tabel hiernaast staat de leeftijdsopbouw van leraren in het primair onderwijs (po) en het voortgezet onderwijs (vo) in de jaren 1995 en 2005.

- Teken staafdiagrammen van de verdeling van de leeftijdsopbouw voor 1995 en 2005 in het po.
- Beschrijf de verschillen tussen beide verdelingen.
Leg met name uit waar je de nieuwe instroom van leraren in het po aan herkent.
- Bepaal de klassenmiddens en geef daarmee een schatting van de gemiddelde leeftijden in het po in 1995 en in 2005.
- Teken staafdiagrammen van de verdeling van de leeftijdsopbouw voor 1995 en 2005 in het vo.
- Beschrijf de verschillen tussen beide verdelingen.
Leg met name uit waaraan je de uitstroom van leraren in het vo kunt zien.
- Teken de cumulatieve relatieve frequentiepolygonen voor het vo in één figuur. Teken er boxplots bij.
- Vergelijk beide verdelingen nog eens. Welke conclusie trek je voor het vo?

Leeftijdsopbouw leraren						
			po		vo	
leeftijd			1995	2005	1995	2005
20 t/m 24			3,7	6,1	0,7	3,3
25 t/m 29			11,9	15,6	5,3	9,9
30 t/m 34			12,3	11,1	8,7	9,7
35 t/m 39			18,2	8,1	14,5	8,8
40 t/m 44			21,3	10,6	21,0	10,9
45 t/m 49			18,3	16,6	22,5	16,3
50 t/m 54			9,5	16,7	17,3	19,4
55 t/m 59			4,4	12,7	9,1	17,1
60 t/m 64			0,5	2,5	1,0	4,5
			100,0	100,0	100,0	100,0

Opgave 72

Gebruik het bestand ►SPORTPRESTATIES.

- Vergelijk de verdelingen voor de prestaties van de jongens en de meisjes bij het vergooien. Gebruik daarbij de verschillen in vorm, de verschillen tussen de centrummaten en de spreiding van de verdelingen. Gebruik hiervoor staafdiagrammen, boxplots en de kentallen.
- Maak cumulatieve relatieve frequentieverdelingen. Probeer ook daaruit conclusies te trekken.

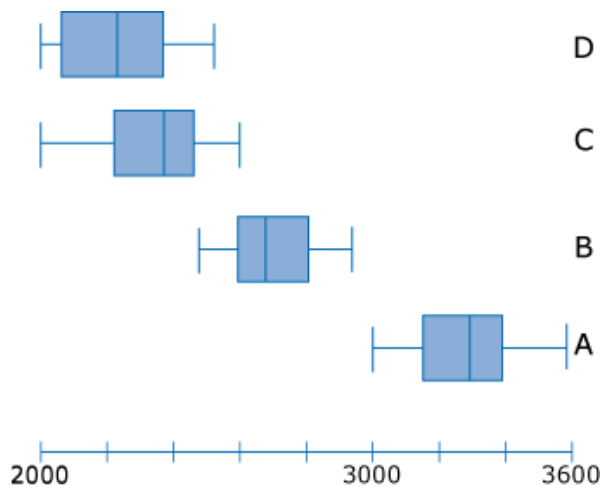
Opgave 73

Gebruik het bestand ►5ATH. We willen weten of er verschil is in de cijfers voor geschiedenis als we de jongens en de meisjes vergelijken.

- Maak twee staafdiagrammen van de relatieve frequenties voor het geschiedeniscijfer voor jongens en voor meisjes. Wat valt je op?
- Wordt dit beeld bevestigd door de boxplot?
- Maak twee relatieve cumulatieve frequentiepolygonen en schrijf een stukje tekst over de verschillen tussen jongens en meisjes voor het geschiedeniscijfer.
Welke kanttekening kun je bij deze tekst maken?

Opgave 74

Hiernaast staan boxplots die het aantal branduren van vier type lampen beschrijven.



Je ziet onmiddellijk dat lampen van type A een langere brandtijd hebben dan die van alle andere types.

- Leg uit hoe je onmiddellijk kunt zien dat lampen van type A een langere brandtijd hebben dan die van type B, C of D.
- Waarom weet je nog steeds niet 100 procent zeker dat elke lamp van type A langer brandt dan een lamp van de andere types?
- Hoeveel procent van de lampen van type C gaat langer mee dan de lamp van type B met de kortste brandtijd?
- Zijn lampen van type C en van type D nog zinvol te vergelijken met behulp van hun boxplots?
Zo ja: doe een uitspraak die je nog veilig kunt doen (als de steekproeven groot genoeg zijn).
Zo nee: leg uit waarom je geen enkele zinvolle uitspraak kunt doen over het verschil in branduren van type C en type D.

Opgave 75

Bekijk de dataset ► **LICHMAAT**. Je vindt daar enkele lichaamsmaten van 404 mannen en vrouwen. Een bepaalde lichaamsmaat wordt verschillend genoemd voor mannen en vrouwen als de boxen van de twee boxplots geen overlap hebben.

- Wat vind je van deze afspraak? Formuleer deze uitspraak met behulp van percentages.
- Vergelijk de boxplots van de handspanne van mannen en vrouwen.
Kun je zeggen dat mannen een grotere handspanne hebben dan vrouwen?
- En hoe zit dat met de lengtes van mannen en vrouwen?
- En hoe zit het met de verschillen tussen mannen en vrouwen bij de variabelen *gewicht* en *schoenmaat*?

§ 2.4.5 Om te onthouden

Verschillen tussen twee groepen kun je beschrijven door te kijken naar:

- Vorm van de verdeling (in het staafdiagram).
- Verschillen in centrummaten (zichtbaar in verschillende representaties).
- Verschillen in spreidingsmaten (zichtbaar in boxplot en cumulatief frequentiepolygoon).

Wees kritisch op je conclusies. Soms zijn er niet genoeg gegevens om een echte conclusie te trekken. Als de groepen niet even groot zijn, dan is het verstandig om percentages te nemen.

§ 2.4.6 Geïntegreerd oefenen

Opgave 76

Verzin een manier om iemands reactiesnelheid te meten.

Maak vervolgens een staafdiagram van de verdeling van zijn reactiesnelheden.

Doe dit voor meerdere personen en zet de gegevens in de computer.

- Maak voor een aantal personen een staafdiagram van de reactiesnelheden.
- Beschrijf de vorm van deze staafdiagrammen.
- Bereken de centrummaten en spreidingsmaten die zinvol zijn en leg uit waarom dat het geval is.
- Vergelijk de prestaties van deze personen en gebruik daarbij het voorgaande.
- Bedenk nu zelf andere klassenindelingen en andere representaties waarmee je de reactiesnelheden nog weer extra met elkaar kunt vergelijken: benoem welke informatie je nu nog extra hebt verkregen.



§ 2.5 Samenhang tussen twee variabelen

§ 2.5.1 Introductie

De verkoop van ijs stijgt als de temperatuur stijgt, terwijl de verkoop van handschoenen stijgt als de temperatuur daalt. In beide gevallen is sprake van een statistisch verband tussen temperatuur en verkoop. In deze paragraaf kijken we naar het verband tussen twee variabelen en welk soort conclusies je kunt trekken wanneer je een verband ziet tussen twee variabelen.

§ 2.5.2 Centrale vraag

Centrale vraag 1

Is er een verband tussen de lengte en het gewicht in de dataset van 154 leerlingen?

Centrale vraag 2

Speelt het geslacht een rol bij de keuze tussen wiskunde A en wiskunde B?

Centrale vraag 3

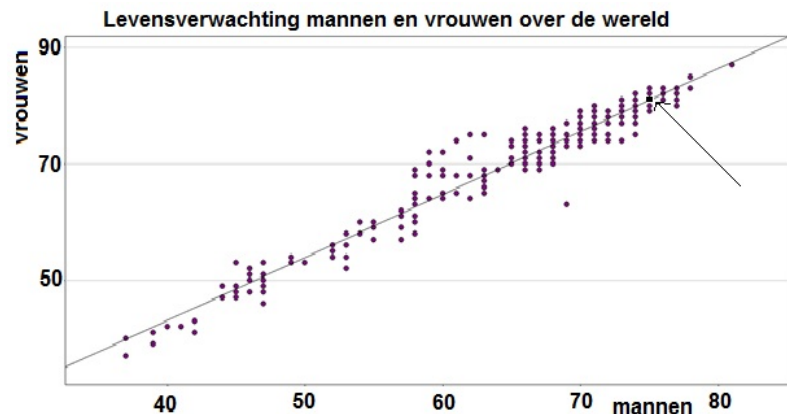
Welke datarepresentaties kun je gebruiken om te kijken of er een verband bestaat tussen twee variabelen?



§ 2.5.3 Samenhang

De volgende grafiek laat zien dat er een heel sterk verband is tussen de levensverwachting van mannen en vrouwen. Iedere stip is een land ergens in de wereld.

De plek van Nederland wordt aangegeven met de pijl. Een dergelijke grafiek heet **spreidingsdiagram**, de verzameling punten in het spreidingsdiagram noemen we **puntenwolk**.



Opgave 77

- Hoe kun je zien dat er een duidelijk verband bestaat tussen de levensverwachting van mannen en van vrouwen?
- Stel dat er een land is dat niet in deze grafiek is opgenomen. De levensverwachting van vrouwen in dit land is 73. Wat verwacht je dat de levensverwachting voor mannen in dat land is?

📊 Opgave 78

Bekijk de dataset ► **GEGEVENS154LEERLINGEN** nog een keer.

- Teken een puntenwolk met horizontaal het gewicht en verticaal de lengte. Vind jij dat er een verband bestaat tussen lengte en gewicht?
- Jan hoort eigenlijk ook bij deze groep leerlingen, maar hij heeft niet meegedaan aan het onderzoek. Hij is 198 centimeter lang. Kun je voorspellen tussen welke waarden zijn gewicht ligt?

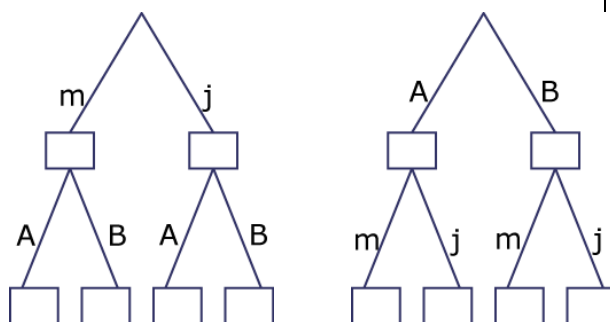


Opgave 79

Hier zie je hoe je in een kruistabel de variabelen *geslacht* (J = jongen, M = meisje) en *wiskundegroep* met elkaar kunt combineren.

- Hoeveel meisjes kiezen wiskunde B?
- Vul in beide boomdiagrammen de aantallen in.

Wiskunde- groep	Geslacht		Totaal
	J	M	
A	13	30	43
B	56	55	111
Totaal	69	85	154

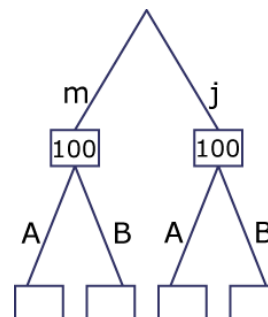


- Welk percentage van de meisjes kiest wiskunde B? Rond af op 1 decimaal nauwkeurig.
 - Hoeveel jongens kiezen wiskunde B? Welke percentage van de jongens is dat?
 - Bereken het percentage wiskunde B-leerlingen dat een jongen is.
- Welk diagram kun je daarvoor het beste gebruiken?

Opgave 80

In het boomdiagram hiernaast zijn de meisjes en de jongens elk op 100 procent gesteld.

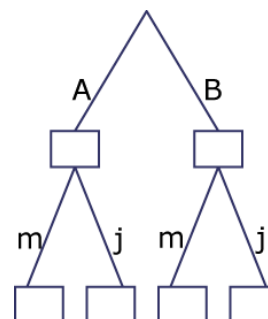
- Vul het boomdiagram verder in.
- Maak een nieuwe kruistabel met de hiervoor berekende percentages.
- Hoeveel verschillen de percentages bij de meisjes?
Wat valt je daarbij op?
- Wat valt je op in vergelijking met de percentages van de jongens?



Opgave 81

Je kunt in de kruistabel van de vorige opgave ook kijken naar het percentage wiskunde B-leerlingen dat meisje dan wel jongen is.

- Hoeveel procent van de leerlingen bij wiskunde B is een meisje?
- Waarom ziet het bijpassende boomdiagram er nu zo uit als dit hiernaast?
- Maak weer de bijpassende kruistabel met percentages.
- Hoeveel verschillen de percentages jongens en meisjes bij de wiskunde B-leerlingen?



Antwoord op centrale vraag 1

Er is een verband tussen lengte en gewicht, al is het zeker niet zo dat een grotere lengte veroorzaakt dat je daardoor automatisch een groter gewicht hebt.

Een statistisch verband is een kenmerk van een hele dataset en niet van een enkele persoon. Je kunt wel met een zekere waarschijnlijkheid een voorspelling doen over het gewicht van een leerling die niet in deze dataset is opgenomen als je de lengte weet.

Antwoord op centrale vraag 2

Antwoord op centrale vraag 3

Je weet dat er kwantitatieve en kwalitatieve variabelen zijn. Wanneer je verband wilt aantonen tussen kwantitatieve variabelen, dan kun je een spreidingsdiagram gebruiken.

Voor kwalitatieve variabelen kun je gebruikmaken van een kruistabel.

Daarbij moet je goed afspreken waar je naar kijkt:

- Verdeling van de variabele *geslacht* over de soort wiskunde.
OF:
- Verdeling van de variabele *wiskundegroep* over het geslacht.

Kijk je naar de verdeling van de variabele *wiskundegroep* over het geslacht, dan worden de totalen per wiskundegroep op 100 procent gesteld. In de kruistabel die je hiervoor tegenkwam betekent dit **horizontaal percenteren**: het totaal van elke rij wordt 100 procent.

Kijk je naar de verdeling van de variabele *geslacht* over de soort wiskunde, dan worden de totalen per geslachtsgroep op 100 procent gesteld. In de kruistabel die je hiervoor tegenkwam betekent dit **verticaal percenteren**: het totaal van elke kolom wordt 100 procent.



§ 2.5.4 Oefenen

De verkoop van ijs stijgt als de temperatuur stijgt. Een toename van temperatuur heeft dus tot gevolg dat de ijsverkoop stijgt. Dit is een **causaal** verband: er is sprake van oorzaak en gevolg.
Het aantal gesignaleerde ooievaars neemt in de lente toe en het aantal geboorten eveneens. Hier is wel sprake van een verband, maar het is geen causaal verband, er worden niet meer baby's geboren *omdat* er meer ooievaars zijn.

Opgave 82

Hier zie je een kruistabel waarin de variabele *geslacht* is uitgezet tegen *plezier* (met plezier naar school gaan). De variabele *plezier* is een vijfpuntsschaal: 1=nee, 2=weinig, 3=neutraal, 4=behoorlijk, 5=veel.

Aantal van plezier	plezier ▼					
geslacht ▼	1	2	3	4	5	Eindtotaal
i	4	13	18	22	13	70
m	10	14	23	24	13	84
Eindtotaal	14	27	41	46	26	154

- Heb je hier met kwalitatieve of kwantitatieve variabelen te maken?
- Is bij de variabele *geslacht* de volgorde van belang? En hoe zit dat bij de variabele *plezier*?

Opmerking

Wanneer je een volgorde kunt vaststellen in een kwalitatieve variabele, dan noem je deze **ordinaal**.
Wanneer je geen volgorde kunt vaststellen is er sprake van een **nominale** variabele.

Opgave 83

Hiernaast zie je de kruistabel van de variabelen *wiskundecijfer* (havo eind 3) en *profiel*.

Aantallen leerlingen	Profiel				
Wiskundecijfer	CM	EM	NG	NT	Totaal
5	1	1	1		3
6	6	6	5	3	20
7		18	23	7	48
8	3	13	22	15	53
9			6	15	21
10		2			2
Totaal	10	40	57	40	147

- Welke van beide variabelen is kwalitatief? Is die variabele ordinaal of nominaal?
- Je wilt nu de frequentieverdeling van de variabele *wiskundecijfer* over de profielen bestuderen. Hoe ga je dan percenteren: horizontaal of verticaal? Licht je antwoord toe.
- Maak nu de bij b passende kruistabel met percentages.
- Kun je door de percentages te vergelijken een conclusie trekken over de verdeling van het *wiskundecijfer* over de profielen?

Opgave 84

Bekijk de kruistabel van de voorgaande opgave nog eens.

Je wilt de verdeling van de variabele *profiel* over de wiskundecijfers havo eind 3 bekijken. Maak een bijpassende kruistabel met percentages en probeer ook nu conclusies te trekken.

Opgave 85

Bekijk de dataset ► **GEGEVENS154LEERLINGEN** nog een keer

- Splits de data op de variabele *geslacht*. Teken nu twee aparte puntenwolken van met horizontaal het gewicht en verticaal de lengte. Je ziet dat de puntenwolk van de jongens in zijn geheel iets boven de puntenwolk van de meisjes ligt.
- Er komt een leerling met een lengte van 180 centimeter.
Welk gewicht vind je aannemelijk wanneer dit een jongen is? En wanneer dit een meisje is?

Opgave 86

Er is onderzoek gedaan naar het favoriete avondje uit onder jongeren, zie de tabel hiernaast.

	Jongens	Meisjes
Film	745	667
Disco	580	370

Maak een kruistabel met relatieve gegevens die de verschillen in voorkeur tussen jongens en meisjes laat zien.

Opgave 87

Bij een onderzoek over kleurenblindheid is 1000 mensen gevraagd of ze een vorm van kleurenblindheid hebben. In totaal werden 600 mannen bevroegd, waarvan er 65 aangaven kleurenblind te zijn.

Van de vrouwen bleken er maar 7 kleurenblind te zijn.

- Maak met deze gegevens een kruistabel van de variabelen *kleurenblindheid* en *geslacht*.
- Welke deel van de kleurenblinden in deze groep is van het mannelijk geslacht?
- Welk deel van de mannen in deze groep is kleurenblind?
- Zou er een relatie bestaan tussen beide variabelen? Zo ja, beschrijf dan die relatie.

Opgave 88

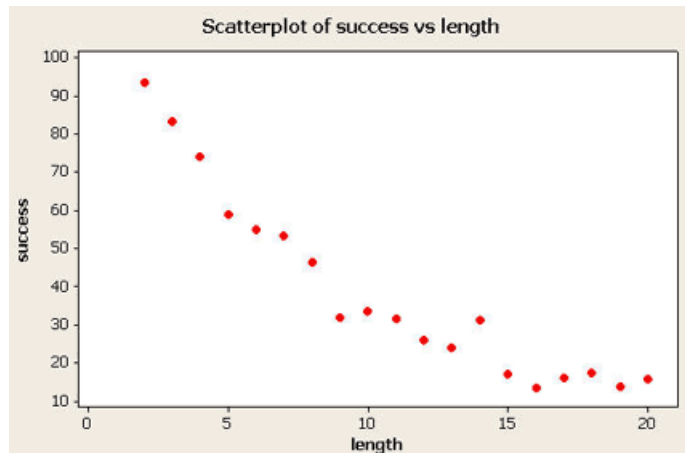
Is er een statistisch verband is tussen ijsverkoop en verkoop van zonnebrillen in de zomer?

Is er sprake van een causaal verband?



Opgave 89

Een groep Amerikaanse golfers heeft bij het putten (het slaan van de golfbal met als doel dat hij in de hole terechtkomt en niet dat hij alleen maar in de buurt van de hole komt) hun percentage successen berekend afhankelijk van de afstand tot de hole (de *length* in m). Hiernaast zie je de resultaten.



- Schat het succespercentage bij een afstand van 15 meter tot de hole.
- Hoe groot is het succespercentage bij 0 meter?
- Er lijkt een statistisch verband te bestaan tussen de variabelen *success* en *length*. Wat kun je zeggen van het succespercentage bij een *length* van 30 meter?
- En wat gebeurt er met het succespercentage als de afstand tot de hole steeds groter wordt?

Opgave 90

Geslacht	Verspringen [cm]					Totaal
	150-199	200 – 249	250-299	300-349	350-399	
Jongen	0	1	7	19	6	33
Meisje	1	5	10	21	3	40
Totaal	1	6	17	40	9	73

- Als je het verspringen tussen jongens en meisjes wilt vergelijken, moet je dan horizontaal of verticaal percentages nemen?
- Maak de tabel die je bij a bedacht hebt en kijk of je een conclusie kunt trekken over het verschil in verspringen tussen jongens en meisjes.
- Maak een puntenwolk bij de variabelen *sprint* en *verspringen*. Onderzoek of je een statistisch verband aantreft tussen beide variabelen en zo ja, beschrijf dan dit verband. Is dit verband duidelijker wanneer je de data splitst op geslacht en twee puntenwolken maakt?

§ 2.5.5 Om te onthouden

Statistisch verband tussen twee variabelen kun je onderzoeken door te kijken naar een **spreidingsdiagram** (puntenwolk) of kruistabel.

Voor kwantitatieve variabelen kun je gebruik maken van een spreidingsdiagram.

Bij kwalitatieve variabelen gebruik je een **kruistabel**.

Kwalitatieve variabelen waarbij je een logische volgorde aan kunt wijzen noem je **ordinaal**. Wanneer er geen logische volgorde is, dan is de variabele **nominaal**.

Met behulp van een puntenwolk of een kruistabel kun je vaststellen of er een statistisch verband is. Er is sprake van een **causaal** verband (oorzakelijk) als er sprake is van oorzaak en gevolg.



§ 2.5.6 Geïntegreerd oefenen

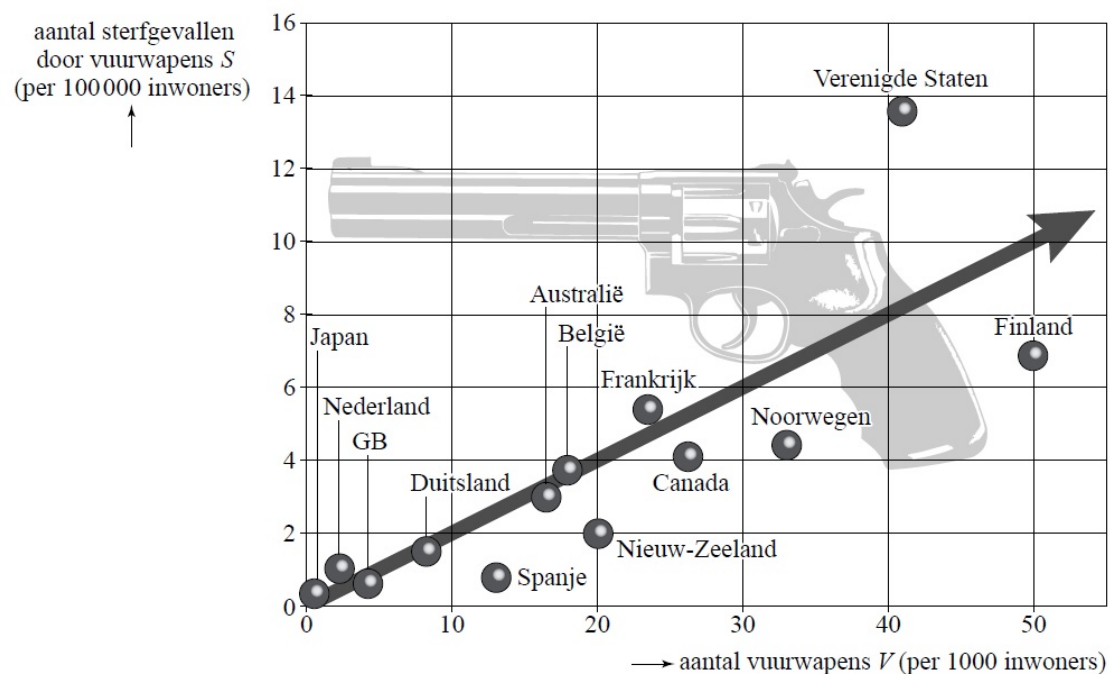
Opgave 91

De regels omtrent het in bezit mogen hebben van vuurwapens zijn per land verschillend. Deze regels staan natuurlijk ook wel eens ter discussie.

Tegenstanders van vuurwapenbezit beweren dat hoe makkelijker mensen aan vuurwapens kunnen komen, hoe meer die gebruikt worden.

Voorstanders van vuurwapenbezit zeggen altijd dat het niet de wapens zijn die doden, maar de mensen. Zij vinden dat mensen vrij moeten zijn om een vuurwapen aan te schaffen, omdat meer vuurwapens niet betekent dat er dan ook meer gebruik van wordt gemaakt.

Het vuurwapenbezit en het aantal dodelijke slachtoffers door vuurwapens is in een aantal landen onderzocht. De onderzoeksresultaten zie je hieronder.

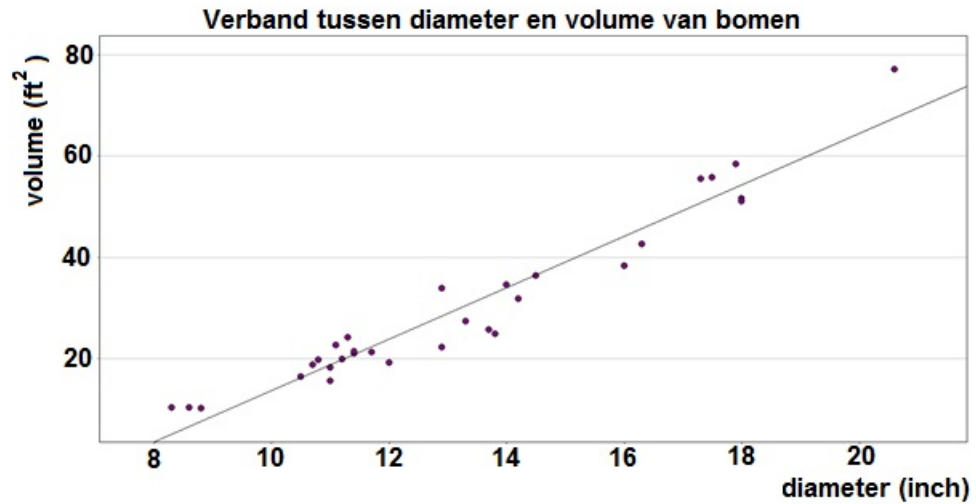


(bron: Cito eindexamen havo A, 2009-I)

- Geef met behulp van de figuur een argument dat voorstanders van vuurwapens kunnen gebruiken.
- Geef met behulp van de figuur een argument dat tegenstanders van vuurwapens kunnen gebruiken.
- Is hier naar jouw mening sprake van een statistisch verband? Van een causaal verband?

Opgave 92

Hieronder zie je een spreidingsdiagram van de diameter en het volume van verschillende bomen.



- Formuleer het causale verband dat uit de bovenstaande grafiek duidelijk wordt.
- Welk volume (in ft^3) verwacht je bij een boom van 30 inch?
- Welk volume (in cm^3) verwacht je bij een boom met een diameter van 40 centimeter?
(1 inch = 2,54 cm, 1 ft = 30,48 cm)

