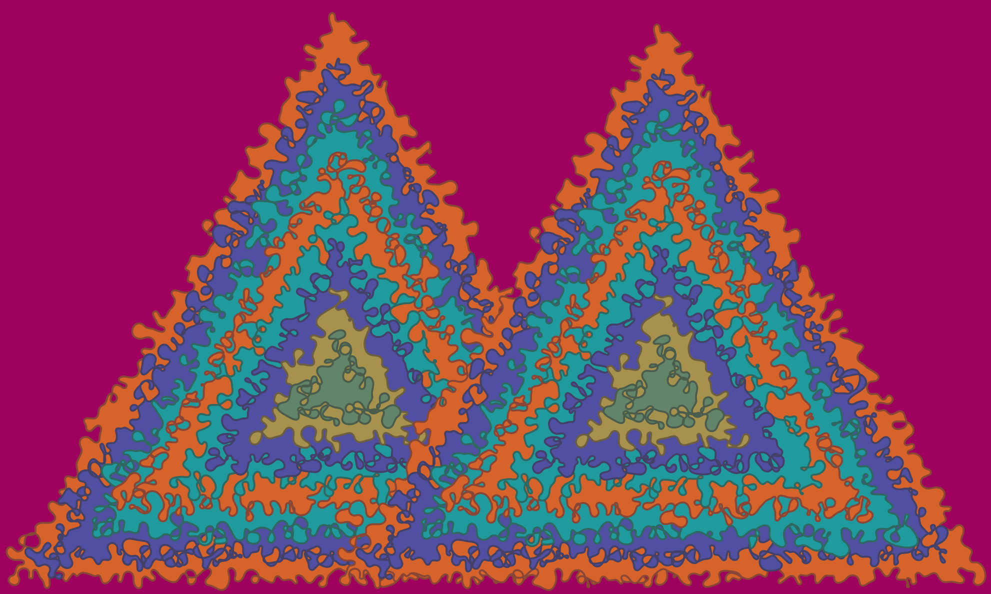
Wiskunde D

4 HAVO

Katern 2

ConTeXt College



Inhoud

Voorwoord

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

Ieder hoofdstuk bestaat uit een aantal paragrafen en wordt steeds afgesloten met een paragraaf *Totaalbeeld* waar de leerstof wordt samengevat en/of herhaald. Iedere paragraaf is ingedeeld in vaste rubrieken die houvast geven bij de bestudering van het lesmateriaal.

- Verkennen
- Uitleg
- Theorie en Voorbeelden
- Verwerken
- Toepassen

Indien er in het lesmateriaal wordt verwezen naar werkbladen dan kun je deze terugvinden op de website en achterin je katern.

1

Kansrekening

1.1 Kansbomen

Inleiding

Bij basketbal wordt per speler het schotpercentage bijgehouden. Als iemand een schotpercentage van 25% heeft, scoort hij bij één op de vier doelpogingen. Je kunt dit percentage daarom opvatten als zijn trefkans bij elke doelpoging.
Maar hoe liggen zijn kansen als hij meerdere doelpogingen doet?

Figuur 1.1

Je leert in dit onderwerp

- kansen bepalen met behulp van een kansboom;
- het vaasmodel met of zonder teruglegging voor het berekenen van kansen.

Voorkennis

- werken met boomdiagrammen;
- kansen berekenen door tellen van mogelijkheden, eventueel met behulp van diagrammen.

Verkennen

Opgave V1

Bij basketbal wordt per speler het schotpercentage bijgehouden. Als iemand een schotpercentage van 25% heeft, scoort hij bij één op de vier doelpogingen. Je kunt dit percentage daarom opvatten als zijn trefkans bij elke doelpoging.
Maar hoe liggen zijn kansen als hij meerdere doelpogingen doet?

a Hoe groot is de kans op twee scores als hij twee doelpogingen doet?

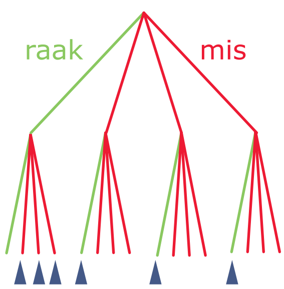
b Hoe groot is de kans op een score als hij twee doelpogingen doet?

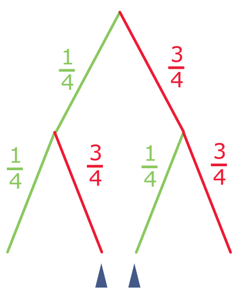
c Hoeveel treffers verwacht je van hem?

Uitleg 1

Bij basketbal wordt per speler het schotpercentage bijgehouden. Als iemand een schotpercentage van 25% heeft, scoort hij bij één op de vier doelpogingen. Je kunt dit percentage daarom opvatten als zijn trefkans bij elke doelpoging.

Om zijn kansen te bepalen bij bijvoorbeeld twee doelpogingen, maak je een boomdiagram. Je ziet één treffer naast drie missers bij elke poging.



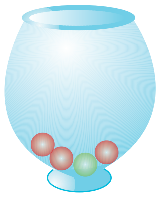


Figuur 1.2

Door missers en treffers samen te voegen, kun je het diagram vereenvoudigen tot een kansboom. Als je de kans wilt berekenen op precies één treffer bij twee doelpogingen, kun je in het boomdiagram de juiste routes tellen: het zijn er zes van de zestien.
In de kansboom moet je dan kansen vermenigvuldigen en optellen:
$\text{P}(X=1) = \frac{1}{4} \cdot \frac{3}{4} + \frac{3}{4} \cdot \frac{1}{4} = \frac{6}{16} = 0{,}375.$

Je kunt ook een andere methode gebruiken om zijn trefkansen te berekenen.

Om zijn kansen te bepalen bij bijvoorbeeld twee doelpogingen kun je dit opvatten als het aselect twee keer trekken van een balletje uit een vaas met één groen balletje (raak) en drie rode balletjes (mis). Je moet dan wel nadat je de eerste keer een balletje hebt getrokken, dit balletje weer terug doen in de vaas en het geheel schudden. Dit is een 'vaasmodel' voor de doelpogingen van de basketballer, en het is 'een trekking met teruglegging'.



Figuur 1.3

Bij elk vaasmodel kun je een kansboom maken om de bijbehorende kansen te berekenen.
Als X het aantal treffers bij twee doelpogingen is, dan geldt ook in het vaasmodel
$\text{P}(X=1) = \frac{1}{4} \cdot \frac{3}{4} + \frac{3}{4} \cdot \frac{1}{4} = \frac{6}{16} = 0{,}375.$

Opgave 1

Ga nu uit van een basketballer met een schotpercentage van 15%.

a Teken een kansboom uitgaande van twee schoten op de basket.

b Bereken de kans op precies één treffer. Geef het antwoord in drie decimalen nauwkeurig.

c Bereken de kans op twee treffers. Geef het antwoord in vier decimalen nauwkeurig.

d Bereken de kans op hoogstens één treffer. Geef het antwoord in vier decimalen nauwkeurig.

Opgave 2

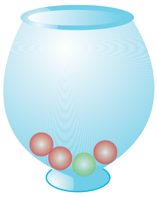
Een basketballer met een schotpercentage van 15% schiet drie keer op de basket.

a Bereken de kans op twee treffers. Geef het antwoord in zes decimalen nauwkeurig.

b Bereken de kans op hoogstens twee treffers. Geef het antwoord in vier decimalen nauwkeurig.

c Bereken de kans op minstens twee treffers. Geef het antwoord in vier decimalen nauwkeurig.

Uitleg 2



Figuur 1.4

In een vaas zitten vier balletjes, drie rode en één groene. Je trekt daaruit aselect drie keer een balletje. De getrokken balletjes leg je niet terug in de vaas ('een trekking zonder teruglegging'). Je wilt weten hoe groot de kans is op eerst een rood balletje, dan een groen balletje en bij de derde trekking weer een rood balletje.

Dit is een vaasmodel met een trekking zonder teruglegging.

Je bent op zoek naar de kans P (rood, groen, rood). Bij de eerste trekking zijn drie van de vier balletjes rood. De kans dat het eerste balletje rood is, is $\frac{3}{4}$. Het eerste getrokken balletje leg je niet terug.

Daarna wil je het groene balletje trekken. Die kans is $\frac{1}{3}$, want er zijn drie balletjes over en één daarvan is groen. Het derde balletje moet weer rood zijn. Er zijn nog twee rode balletjes. De kans op een rood balletje is $\frac{2}{2}$.

P (rood, groen, rood) $= \frac{3}{4} \cdot \frac{1}{3} \cdot \frac{2}{2} = \frac{6}{24} = \frac{1}{4} = 0{,}25$

Opgave 3

Je hebt trekking met terugleggen en trekking zonder terugleggen, zie **Uitleg 2**.

a Er wordt bij de basketballer uit **Uitleg 1** aangenomen dat hij tijdens de schoten op de basket een vast schotpercentage van 25% heeft. Gaat het dan om trekking met of trekking zonder terugleggen? Licht je antwoord toe.

b Waarom kan hier slechts van een aanname sprake zijn? Hoe zit het in werkelijkheid met schotpercentages?

Opgave 4

In een vaas zitten zes balletjes, twee rode en vier groene. Je trekt daaruit aselect en met terugleggen twee keer een balletje.

a Laat in een kansboom alle mogelijkheden zien.

b Hoe groot is de kans op eerst een groen en dan een rood balletje?

c Hoe groot is de kans op twee balletjes van verschillende kleur?

Opgave 5

Uit de vaas met zes balletjes (twee rode en vier groene) trek je nu twee balletjes tegelijk. Je kunt dit zien als het trekken van twee keer één balletje zonder terugleggen.

a Maak een kansboom voor de kleuren.

b Hoe groot is de kans op een rood en een groen balletje?

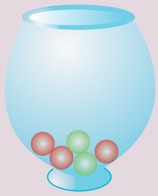
c Hoe groot is de kans op twee balletjes van dezelfde kleur?

Theorie en voorbeelden

Om te onthouden

Veel situaties waarin kansen een rol spelen, kun je beschrijven door middel van aselecte trekking van balletjes uit een vaas. Dit noem je een **vaasmodel** voor de situatie.

Bij elk vaasmodel kun je een **kansboom** maken om kansen te berekenen. Daarbij moet je goed onderscheiden:



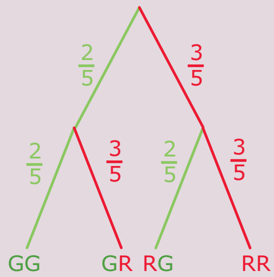
Figuur 1.5

Trekking met teruglegging
Na elke trekking van één balletje doe je dit terug in de vaas en je schudt het geheel. Dan pas trek je een volgend balletje. De kansboom bij een vaas met twee groene en drie rode balletjes zie je hiernaast.

$P(2 \text{ groen}) = P(GG) = \frac{2}{5} \cdot \frac{2}{5}$

$P(\text{eerst groen en daarna rood}) = P(GR) = \frac{2}{5} \cdot \frac{3}{5}$

$P(\text{groen en rood}) = P(GR \text{ of } RG) = \frac{2}{5} \cdot \frac{3}{5} + \frac{3}{5} \cdot \frac{2}{5}$



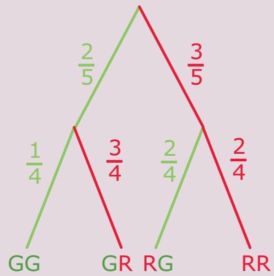
Figuur 1.6

Trekking zonder teruglegging
Na elke trekking van één balletje doe je dit niet terug in de vaas. Je trekt meteen het volgende balletje. (Twee in één greep kan ook.) De kansboom zie je hiernaast.

$P(2 \text{ groen}) = P(GG) = \frac{2}{5} \cdot \frac{1}{4}$

$P(\text{eerst groen en daarna rood}) = P(GR) = \frac{2}{5} \cdot \frac{3}{4}$

$P(\text{groen en rood}) = P(GR \text{ of } RG) = \frac{2}{5} \cdot \frac{3}{4} + \frac{3}{5} \cdot \frac{2}{4}$



Figuur 1.7

In een kansboom vermenigvuldig je de kansen bij een bepaalde route steeds vanaf het beginpunt van de boom. Omdat hier twee balletjes worden getrokken, heeft de kansboom twee ‘lagen’. Het aantal lagen is gelijk aan het aantal getrokken balletjes.

Voorbeeld 1

In een groep van vier mannen en vijf vrouwen worden door loten drie taken (wassen, afwassen en auto wassen) verdeeld.
Als één persoon meerdere van die drie taken mag doen, hoe groot is dan de kans dat er twee taken door een man en één door een vrouw worden uitgevoerd?

Antwoord

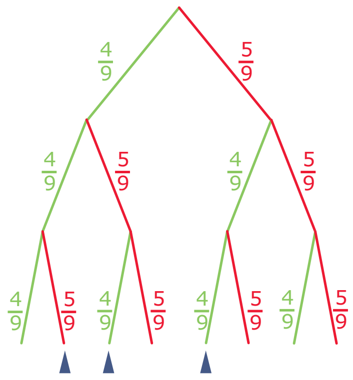
Bijpassend vaasmodel:

- vaas met negen balletjes, vier groen (mannen) en vijf rood (vrouwen);
- aselecte trekking van drie balletjes (aantal taken);
- met teruglegging (want elke persoon mag meerdere taken doen).

Daarbij hoort deze kansboom.

De routes waarbij twee taken door een man en één door een vrouw worden gedaan, zijn aangegeven.
De gevraagde kans is:

P (mmv of mvm of vmm) $= \frac{4}{9} \cdot \frac{4}{9} \cdot \frac{5}{9} + \frac{4}{9} \cdot \frac{5}{9} \cdot \frac{4}{9} + \frac{5}{9} \cdot \frac{4}{9} \cdot \frac{4}{9} = \frac{240}{729} = \frac{80}{243}$.



Figuur 1.8

Opgave 6

Bekijk **Voorbeeld 1**. Er worden kansen berekend bij het door loting verdelen van drie taken onder vier mannen en vijf vrouwen.

a Waaruit blijkt dat het in dit voorbeeld gaat om trekking met terug-leggen?

b Bereken de kans dat er twee taken door een vrouw en één door een man worden gedaan.

c Bereken de kans dat alle taken door een man worden gedaan.

d Bereken de kans dat er hoogstens twee vrouwen één van de drie taken moeten doen.

Opgave 7

Uit een vaas met twee groene en vier rode balletjes pak je tegelijk twee balletjes zonder te kijken. Je wilt de kans op rood en groen balletje berekenen.

a Leg uit waarom die kans $\frac{8}{30} + \frac{8}{30} = \frac{8}{15}$ is.

Je kunt dit ook heel anders aanpakken.

b Ga na dat uit zes balletjes vijftien verschillende paren balletjes zijn te kiezen.

c Bij hoeveel van die paren zijn de balletjes van verschillende kleur? Laat zien dat je nu dezelfde kans krijgt

Voorbeeld 2

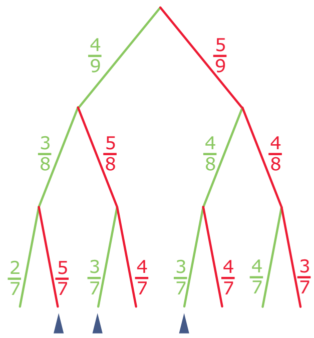
In een groep van vier mannen en vijf vrouwen worden door loten drie taken (wassen, afwassen en auto wassen) verdeeld.
Als elk van die drie taken door een andere persoon moet worden gedaan, hoe groot is dan de kans dat er twee taken door een man en één door een vrouw worden uitgevoerd?

Antwoord

Bijpassend vaasmodel:

- vaas met negen balletjes, vier groen (mannen) en vijf rood (vrouwen);
- aselecte trekking van drie balletjes (aantal taken);
- zonder teruglegging (want elke persoon mag één taak doen en niet meer).

Daarbij hoort deze kansboom.



Figuur 1.9

De routes waarbij twee taken door een man en één taak door een vrouw worden gedaan, zijn aangegeven.
De gevraagde kans is:

P (mmv of mvm of vmm) $= \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{5}{7} + \frac{4}{9} \cdot \frac{5}{8} \cdot \frac{3}{7} + \frac{5}{9} \cdot \frac{4}{8} \cdot \frac{3}{7} = \frac{180}{504} = \frac{5}{14}$

Opgave 8

Bekijk **Voorbeeld 2**. Daarin worden kansen berekend bij het door loting verdelen van drie taken onder vier mannen en vijf vrouwen.

a Waaruit blijkt dat het in dit voorbeeld gaat om trekking zonder terugleggen?

b Bereken de kans dat er twee taken door een vrouw en één door een man worden gedaan.

c Bereken de kans dat alle taken door een man worden gedaan.

d Bereken de kans dat er hoogstens twee vrouwen één van de drie taken moeten doen.

Voorbeeld 3

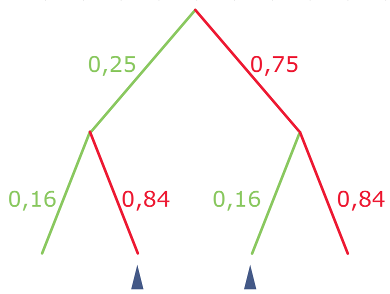
Twee basketballers hebben een verschillend schotpercentage: A heeft een schotpercentage van 25% en B heeft een schotpercentage van 16%.
Beiden doen een doelpoging. Hoe groot is de kans op één treffer?

Antwoord

Voor een vaasmodel van deze situatie heb je nu twee vazen nodig:

- voor A: een vaas met 100 balletjes, 25 groene (treffer) en 75 rode (misser)
  voor B: een vaas met 100 balletjes, 16 groene (treffer) en 84 rode (misser);
- aselecte trekking van één balletje uit elke vaas;
- teruglegging is nu niet relevant, want je trekt maar één balletje uit elke vaas.

Daarbij hoort deze kansboom.



Figuur 1.10

De routes waarbij precies één keer wordt gescoord, zijn aangegeven.
Als X het aantal treffers is, dan is de gevraagde kans:
$\text{P}(X = 1) = 0{,}25 \cdot 0{,}84 + 0{,}75 \cdot 0{,}16 = 0{,}33$.

Opgave 9

In **Voorbeeld 3** gaat het om kansen bij twee basketballers met een verschillend schotpercentage. Ze schieten elk één keer op de basket.

a Hoe groot is de kans op twee treffers?

b Hoe groot is de kans op geen enkele treffer?

c Hoe groot is de kans op minstens één treffer?

Verwerken

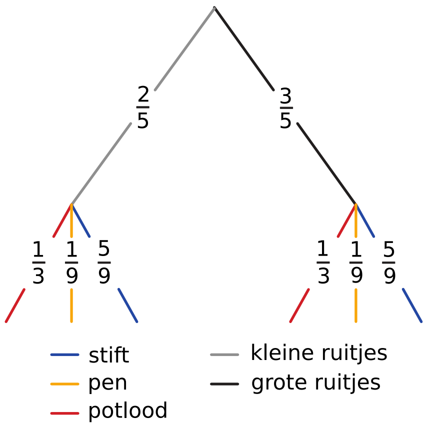
Opgave 10

Bekijk het boomdiagram.

Bas heeft in zijn tas vijf uiterlijk dezelfde schriften, twee met kleine ruitjes en drie met grote ruiten. En hij heeft een etui met vijf stiften, één pen en drie potloden die allemaal dezelfde vorm hebben.

Bas pakt zonder te kijken een schrift en iets om mee te schrijven uit zijn tas.

Hoe groot is de kans dat dit een schrift met grote ruitjes en een stift zijn?



Figuur 1.11

Opgave 11

Joes en z'n zus Wieke mogen ieder een cadeautje uit een grabbelton pakken. In de grabbelton zitten acht grote cadeaus en drie kleine. Wieke pakt als eerste een cadeau en Joes daarna.

a Maak een kansboom bij deze situatie.

b Hoe groot is de kans dat Joes en Wieke beiden een klein cadeau hebben gepakt?

c Hoe groot is de kans dat Wieke een groot cadeau en Joes een klein cadeau heeft gepakt?

d Hoe groot is de kans dat Wieke en Joes samen een klein en een groot cadeau hebben gepakt?

Opgave 12

Arsenal speelt een thuiswedstrijd tegen Juventus. 50% van de bezoekers denkt dat Arsenal wint en $\frac{1}{3}$ deel denkt dat Juventus wint.

Bij de returnwedstrijd rekent $\frac{1}{3}$ deel op winst voor Arsenal en $\frac{1}{3}$ op winst voor Juventus.

a Maak een kansboom voor beide wedstrijden.

b Hoe groot is de kans dat elk van beide teams één van de wedstrijden wint?

Opgave 13

Er wordt met drie dobbelstenen geworpen. Een kansboom kan nu erg groot worden. Misschien heb je er maar een stukje van nodig, of kun je een vaas in gedachten nemen?

a Hoe groot is de kans dat je zeventien of achttien gooit?

b Hoe groot is de kans dat je zestien gooit?

c Hoe groot is de kans dat je minstens twee zessen gooit?

d Voor de vraag naar het aantal zessen kun je een vaasmodel maken. Hoeveel kleuren gebruik je? Hoeveel balletjes van elke kleur heb je nodig?

Opgave 14

In een vaas zitten tien balletjes, zes van hout en vier van plastic. Van de houten balletjes zijn er vier rood en twee blauw. Van de plastic balletjes zijn er drie rood en is er één blauw. Op gevoel zijn hout en plastic niet te onderscheiden. Je trekt twee balletjes uit de vaas. Het gaat om de kleur èn het materiaal van de getrokken balletjes. Neem aan dat het eerst getrokken balletje wordt teruggelegd.

a Bereken de kans dat je eerst een rood houten, dan een blauw plastic balletje trekt.

b Bereken de kans dat je een rood houten en een blauw plastic balletje trekt.

Neem nu aan dat het eerst getrokken balletje niet wordt teruggelegd.

c Bereken de kans dat je eerst een rood houten, dan een blauw plastic balletje trekt.

d Bereken de kans dat je een rood houten en een blauw plastic balletje trekt.

e Als het alleen om de kleur van de twee getrokken balletjes gaat, voldoet een kleinere kansboom. Teken die kansboom voor de gevallen met en zonder terugleggen.

f Bereken in elk van de twee gevallen de kans op twee verschillend gekleurde balletjes.

g De kans op twee verschillend gekleurde balletjes uit de vorige vraag is het grootst als je niet teruglegt. Verklaar dat.

Opgave 15

Er zijn twee taken te doen. Uit een groep van drie mannen en vijf vrouwen moeten twee personen worden geloot die de taken moeten uitvoeren.

Iemand zegt: ‘De kans dat de tweede taak door een man wordt verricht, is gelijk aan de kans dat de eerste taak door een man wordt verricht, want je kunt net zo goed eerst voor de tweede taak loten.’

Bereken of hij gelijk heeft.

Toepassen

Opgave 16: Gomballen

De Grutter verkoopt zakjes met 16 gomballen. De bedoeling is dat in elk zakje 8 rode en 8 gele gomballen zitten. De vulmachine is niet zo precies, zodat dat maar voor de helft van de zakjes klopt. In $\frac{1}{4}$ deel van de zakjes zitten 9 rode en 7 gele gomballen en in het resterende $\frac{1}{4}$ deel van de zakjes zitten 7 rode en 9 gele.

a Je kiest aselect zo'n zakje en daaruit aselect een gombal. Je denkt: rood en geel hebben dezelfde kans, dus de kans dat mijn gombal rood is, is 50%. Leg uit, dat dit inderdaad klopt.

b Je hebt zo’n zakje met 9 rode en 7 gele gomballen. Je eet achter elkaar 3 gomballen, willekeurig gepakt. Is de kans dat ze alle drie geel zijn groter dan 5%?

c Terwijl jij snoept, pikt iemand anders een rode gombal weg. De kans op 3 gele zal dan wel wat groter zijn. Ga na wanneer die het grootst is: als de rode gepikt wordt nadat jij de eerste op hebt of als dat na jouw tweede gombal gebeurt.

Testen

Opgave 17

In een vaas zitten 15 balletjes, 4 witte en 5 rode en 6 blauwe. Er wordt aselect met terugleggen drie keer een balletje getrokken.

a Bereken de kans dat er drie keer een rood balletje wordt getrokken.

b Bereken de kans dat er twee rode balletjes worden getrokken.

c Bereken de kans dat alle balletjes een andere kleur hebben.

Opgave 18

In een vaas zitten 15 balletjes, 4 witte en 5 rode en 6 blauwe. Er wordt aselect drie balletjes in één keer getrokken.

a Bereken de kans dat er drie rode balletjes worden getrokken.

b Bereken de kans dat er twee rode en één wit balletje worden getrokken.

c Bereken de kans dat alle balletjes een andere kleur hebben.

Opgave 19

Ongeveer 1 op de 12 Nederlandse mannen is kleurenblind. Aan hun uiterlijk kun je dit niet zien. Een politieagent houdt een auto aan die door een rood stoplicht reed. Er blijken vier mannen in te zitten. Hij vraagt zich af hoe groot de kans is dat twee inzittenden kleurenblind zijn.

a Maak een bijpassende kansboom.

b Beantwoord de vraag van de politieagent. Geef het antwoord in vier decimalen nauwkeurig.

1.2 Kansen optellen en aftrekken

Inleiding

Je hebt nu het vaasmodel en de bijbehorende kansboom leren kennen. Het lijkt een handig instrument om alle kansproblemen op te lossen, maar toch is het in de praktijk niet altijd even bruikbaar. Zodra het om grotere aantallen trekkingen gaat, wordt een kansboom onoverzichtelijk. Bij trekkingen uit een kaartspel bijvoorbeeld is dit al snel het geval. Dan is het handiger om terug te vallen op regels die beschrijven wanneer je kansen moet optellen en wanneer je ze juist moet vermenigvuldigen.
Daarom wordt nu de kansrekening iets exacter opgebouwd: de kansrekening is een stelsel van regels voor het rekenen met kansen.

Je leert in dit onderwerp

- kansen berekenen met behulp van de somregel en de complementregel;
- de basisbegrippen en de basisregels van de kansrekening;
- een venndiagram aflezen en bijbehorende kansen berekenen.

Voorkennis

- kansen bepalen met behulp van een kansboom;
- het vaasmodel met of zonder teruglegging voor het berekenen van kansen;
- een venndiagram maken.

Verkennen

Opgave V1

Een volledig kaartspel kent 52 kaarten, van elke ‘kleur’ evenveel.
Je trekt aselect uit zo'n kaartspel één kaart.

a Bereken de kans op hartenaas.

b Bereken de kans op hartentwaalf.

c Bereken de kans op een hartenkaart.

d Bereken de kans op geen aas.

e Bereken de kans op een hartenkaart of een ruitenkaart.

f Bereken de kans op een hartenkaart of een boer.

g Bereken de kans op een hartenkaart of een plaatje.

Uitleg

Een volledig kaartspel kent 52 kaarten, van elke 'kleur' (harten, klaveren, ruiten of schoppen) evenveel. Een plaatje is een aas, heer, vrouw of boer.
Je trekt aselect uit zo'n kaartspel één kaart.

- De kans op een hartenkaart is:

  $P(\text{hartenkaart}) = \frac{13}{52}$ want daar zijn er 13 van.
- De kans op hartentwaalf is:
  $P(\text{hartentwaalf}) = 0$ want zo'n kaart bestaat niet.
- De kans op geen hartenkaart is:

  $P(\text{geen hartenkaart}) = \frac{52-13}{52} = \frac{52}{52} - \frac{13}{52} = 1 - \frac{13}{52}$.
- De kans op een hartenkaart of een ruitenkaart is:

  $P(\text{harten of ruiten}) = \frac{13+13}{52} = \frac{13}{52} + \frac{13}{52} = \frac{26}{52} = \frac{1}{2}$.
- De kans op een boer, een aas of een hartenvrouw is:

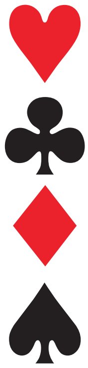
  $P(\text{boer of aas of hartenvrouw}) = \frac{4+4+1}{52} = \frac{4}{52} + \frac{4}{52} + \frac{1}{52} = \frac{9}{52}$.

Het lijkt erop dat je bij 'of' eenvoudigweg de kansen kunt optellen. Maar dat is hier zo, omdat de mogelijkheden elkaar 'wederzijds uitsluiten'. Vraag je namelijk naar een hartenkaart of een boer, dan zijn er niet $13 + 4$ gunstige mogelijkheden, maar slechts $13 + 4 - 1$ vanwege de hartenboer die anders twee keer wordt geteld.

'Hartenkaart' en 'boer' sluiten elkaar niet wederzijds uit.

- De kans op een hartenkaart of boer is:

  $P(\text{hartenkaart of boer}) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$.



harten
aas, heer, vrouw, boer
10, 9, 8, 7, 6, 5, 4, 3, 2

klaveren
aas, heer, vrouw, boer
10, 9, 8, 7, 6, 5, 4, 3, 2

ruiten
aas, heer, vrouw, boer
10, 9, 8, 7, 6, 5, 4, 3, 2

schoppen
aas, heer, vrouw, boer
10, 9, 8, 7, 6, 5, 4, 3, 2

Figuur 2.1

Opgave 1

Lees de **Uitleg** goed door.
Uit een compleet spel speelkaarten wordt aselect een kaart getrokken.

a Hoe groot is de kans dat het een plaatje is?

b Hoe groot is de kans dat het geen plaatje is?

c Hoe groot is de kans dat het een schoppenkaart is?

d Hoe groot is de kans dat het een schoppenplaatje is?

e Hoe groot is de kans dat het een schoppenkaart is of een heer?

f Waarom kun je bij e niet gewoon de kans op een schoppenkaart en de kans op een heer optellen?

g Wordt de kans op schoppen of heer kleiner als ruitenheer en schoppenaas in het spel ontbreken?

Theorie en voorbeelden

Om te onthouden

Bij een **kansexperiment**, zoals het trekken van een kaart uit een kaartspel, bestaat een **uitkomstenverzameling** (in dit geval met 52 mogelijkheden). Een **gebeurtenis**, zoals 'het trekken van een tien', is dan een deel van die uitkomstenverzameling (er zijn vier tienen). Bij elke gebeurtenis hoort een bepaalde **kans**. Hierbij gelden de volgende kansregels:

- De kans op een **onmogelijke gebeurtenis** (niets uit de uitkomstenverzameling) is 0.
- De kans op een **zekere gebeurtenis** (de complete uitkomstenverzameling) is 1.
- Is G een gebeurtenis, dan is $\text{P}(G)$ de kans op die gebeurtenis en $0 \leq \text{P}(G) \leq 1$.
- De **complementregel**:
  Is niet-G de ontkenning van gebeurtenis G, dan is P (niet-G) $= 1 - \text{P}(G)$.
  Je noemt niet-G en G wel complementaire gebeurtenissen.
- De **somregel**:
  - als de gebeurtenissen G_1 en G_2 elkaar **wederzijds uitsluiten**, dan is
    $\text{P}(G_1 \text{ of } G_2) = \text{P}(G_1) + \text{P}(G_2)$.
  - als de gebeurtenissen G_1 en G_2 elkaar **niet wederzijds uitsluiten**, dan is
    $\text{P}(G_1 \text{ of } G_2) = \text{P}(G_1) + \text{P}(G_2) - \text{P}(G_1 \text{ en } G_2)$.

Je ziet dat voor twee gebeurtenissen G_1 en G_2 die elkaar wederzijds uitsluiten, geldt: $\text{P}(G_1 \text{ en } G_2) = 0$.
Een **venndiagram** is een grafische weergave. Daarin worden de kansen of aantallen van maximaal drie gebeurtenissen of eigenschappen die wel of niet van toepassing zijn, weergegeven.

Voorbeeld 1

Een volledig kaartspel kent 52 kaarten, van elke 'kleur' evenveel. Je trekt aselect uit zo'n kaartspel één kaart. Bereken de kans op een hartenkaart of een plaatje.

Antwoord

Er zijn dertien hartenkaarten en $4 \cdot 4 = 16$ plaatjes.
Maar deze gebeurtenissen sluiten elkaar niet uit: er zijn vier hartenplaatjes.

De gevraagde kans is dus:

$\text{P}(\text{hartenkaart of plaatje}) = \frac{13}{52} + \frac{16}{52} - \frac{4}{52} = \frac{25}{52}$.

Opgave 2

Bekijk het kaartspel nog eens, zie ook **Voorbeeld 1**. Je trekt er aselect één kaart uit.

Welke van de volgende gebeurtenissen sluiten elkaar uit?

A. Harten kaart of schoppen kaart.
B. Harten kaart of vrouw.
C. Kaart met even getal of plaatje.
D. Kaart met even getal of ruiten kaart.

Opgave 3

Je trekt aselect één kaart uit een kaartspel. Sluiten de volgende gebeurtenissen elkaar uit? Bereken de bijbehorende kans.

a Hartenkaart of schoppenkaart.

b Hartenkaart of vrouw.

c Kaart met even getal of plaatje.

d Kaart met even getal of ruitenkaart.

Voorbeeld 2

Je trekt een lot uit een serie loten met de nummers 10, 11, 12, ..., 99.
Heb je één 2 of één 3 in het lotnummer, dan heb je prijs.
Hoe groot is de kans hierop?

Antwoord

Eerst het cijfer 2:

De kans op links een 2 is $\frac{10}{90}$ en de kans op rechts een 2 is $\frac{9}{90}$.

Dus de kans op één 2 is (sluit 22 uit):

$\frac{10}{90} + \frac{9}{90} - \frac{1}{90} = \frac{18}{90}$

De kans op één 3 in het lotnummer is op dezelfde manier $\frac{18}{90}$.

De kans op een 2 of een 3 in het lotnummer is (sluit 23 en 32 uit):
$\frac{18}{90} + \frac{18}{90} - \frac{2}{90} = \frac{34}{90} = \frac{17}{45}$.

Opgave 4

In **Voorbeeld 2** gaat het om de trekking bij een loterij.

a Hoe groot is de kans dat het getrokken briefje het cijfer 0 bevat?

b Hoe groot is de kans dat het getrokken briefje het cijfer 0 en het cijfer 2 bevat?

c Hoe groot is de kans dat het getrokken briefje het cijfer 0 of het cijfer 2 bevat?

d Bereken de kans dat het getrokken briefje geen 0 en ook geen 2 bevat.

Maak gebruik van de complementregel. Je hebt bij c immers uitgerekend hoe groot de kans is op een briefje met een 0 of een 2.

Opgave 5

Je gooit met twee gewone dobbelstenen, een rode en een groene. R is het aantal ogen op de rode dobbelsteen, W het aantal ogen op de groene dobbelsteen.

a Maak een overzicht van alle mogelijkheden.

b Hoe groot is $\text{P}(R = 5)$?

c Hoe groot is $\text{P}(W = 4)$?

d Hoe groot is $\text{P}(R = 5 \text{ en } W = 4)$?

e Sluiten de gebeurtenissen $R = 5$ en $W = 4$ elkaar wederzijds uit?

f Hoe groot is $\text{P}(R = 5 \text{ of } W = 4)$?

Voorbeeld 3

Je mag vier keer met een dobbelsteen gooien. Maar je gooit tot je een zes gooit. Dan stop je. Hoe groot is de kans dat je met maximaal vier keer gooien één keer zes gooit?

Antwoord

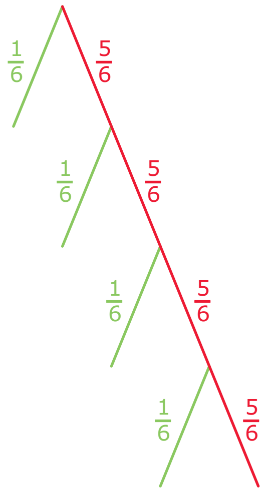
Maak hierbij een kansboom: groen stelt het werpen van een zes voor en rood stelt geen zes voor. De kansen zijn:

- meteen een zes gooien: kans $\frac{1}{6}$
- pas bij de tweede worp een zes gooien: kans $\frac{5}{6} \cdot \frac{1}{6}$
- pas bij de derde worp een zes gooien: kans $\left(\frac{5}{6}\right)^2 \cdot \frac{1}{6}$
- pas bij de vierde worp een zes gooien: kans $\left(\frac{5}{6}\right)^3 \cdot \frac{1}{6}$

Omdat deze vier gevallen elkaar uitsluiten, mag je de kansen optellen.

Dit kan echter eenvoudiger door vast te stellen dat de complementaire gebeurtenis is: vier keer achter elkaar geen zes gooien. Daarbij hoort een kans van $\left(\frac{5}{6}\right)^4$.

De gevraagde kans is daarom $1 - \left(\frac{5}{6}\right)^4 \approx 0{,}52$.



Figuur 2.2

Opgave 6

In **Voorbeeld 3** gaat het om het werpen met een dobbelsteen tot je een zes gooit. Stel je mag tien keer proberen.

a Hoe groot is de kans dat dit lukt? Geef het antwoord in drie decimalen nauwkeurig.

b Hoe groot is de kans dat je bij de derde worp voor het eerst een zes gooit?

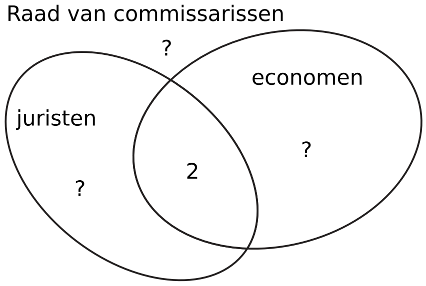
c Hoe groot is de kans dat je bij de achtste worp voor het eerst een zes gooit?

d Hoe groot is de kans dat je in de tien worpen een zes gooit?

e Waarom is de complementregel nu erg handig?

Opgave 7

De raad van commissarissen van een bouwonderneming heeft elf leden, onder wie vijf economen en vier juristen. Twee van de economen zijn ook jurist. De leden zijn om de beurt een maand voorzitter. De volgorde is door loten vastgesteld. Elke maand wordt weer opnieuw geloot, waardoor iemand ook twee maanden achter elkaar voorzitter zou kunnen zijn.



Figuur 2.3

a Vul het venndiagram verder in.

b Hoe groot is de kans dat de voorzitter deze maand econoom en jurist is?

c Hoe groot is de kans dat de voorzitter deze maand econoom of jurist is?

d Hoe groot is de kans dat zowel deze maand als de volgende maand de voorzitter econoom of jurist is?

Verwerken

Opgave 8

In een klas van 31 leerlingen zitten:

- 10 jongens die vijftien jaar zijn;
- 8 jongens die zestien jaar zijn;
- 7 meisjes die vijftien jaar zijn;
- 6 meisjes die zestien jaar zijn.

a Sluiten de gebeurtenissen ‘meisje’ en ‘jongen’ elkaar uit?

Willekeurig kiest de docent een leerling uit deze klas.

b Bereken de kans dat dit een meisje of een jongen is.

c Bereken de kans dat dit een vijftienjarig meisje of een vijftienjarige jongen is.

Opgave 9

Bij een bloemenkraampje zijn nog 25 rozen en 20 tulpen te koop:

- 10 witte tulpen
- 5 gele tulpen
- 5 paarse tulpen
- 12 witte rozen
- 13 gele rozen

De verkoper pakt, zonder te kijken, een bloem.

a Hoe groot is de kans op een roos?

b Hoe groot is de kans op een paarse bloem?

c Hoe groot is de kans op géén paarse bloem?

d Hoe groot is de kans op een gele bloem?

e Hoe groot is de kans op een gele bloem of een tulp?

Opgave 10

Een spel kaarten bevat van elk van de vier ‘kleuren’ alleen de kaarten 7, 8, 9, 10, boer, vrouw, heer en aas. Totaal 32 kaarten. Beantwoord de vragen zowel door tellen van gunstige mogelijkheden als door gebruik van de somregel.

a Hoe groot is de kans dat een uit zo'n spel getrokken kaart een ruiten of een plaatje is?

b Hoe groot is de kans dat een uit zo'n spel getrokken kaart een harten of een 9 of een 10 is?

c Hoe groot is de kans dat een uit zo'n spel getrokken kaart een 9 of een 10 is of geen harten?

Opgave 11

Voor de ontwikkeling van kinderen zijn doosjes in de handel gebracht met plastic rondjes, vierkantjes, rechthoekjes en driehoekjes. Van elke soort zijn er grote en kleine stukjes. Van elke soort en elke grootte zijn er twee rode stukjes, twee gele en twee blauwe. In totaal zijn er dus 48 stuks.
Bereken voor een aselect gekozen stukje de kans dat:

a Het stukje geel of een vierkantje is.

b Het stukje is rood of geen vierkantje.

c Het stukje is klein of geen vierkantje.

d Het stukje is blauw of geel of een driehoekje.

Opgave 12

Bij een spel moet je eerst kruis of munt gooien. Gooi je kruis, dan mag je met één dobbelsteen gooien, gooi je munt, dan mag je met twee dobbelstenen gooien. Bereken de volgende kansen.

a De kans dat je twaalf ogen gooit.

b De kans dat je zeven ogen gooit.

c De kans dat je zeven of twaalf ogen gooit.

d De kans dat je meer of minder dan zeven ogen gooit.

e De kans dat je zes ogen gooit.

Opgave 13

Van de leerlingen van een school is 52% meisje, de rest jongen. Een op de dertien meisjes draagt een hoofddoek, een op de zestien jongens draagt een basketbalpet.

a Hoe groot is de kans dat een aselect aangewezen leerling een basketbalpet draagt?

b Hoe groot is de kans dat een aselect aangewezen leerling een meisje is of iets op het hoofd draagt?

c Hoe groot is de kans dat een aselect aangewezen leerling een jongen is of niets op het hoofd draagt?

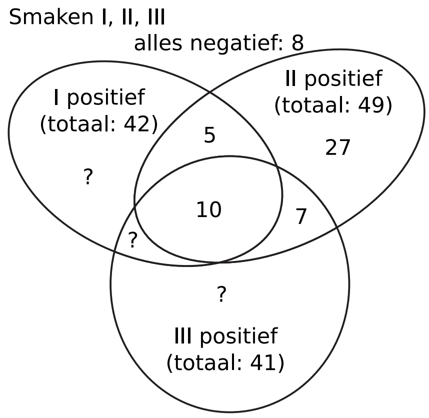
d Wat is de complementaire gebeurtenis van die bij c?

Toepassen

Opgave 14: Het kaakje

Een fabrikant van koekjes wil een nieuwe variant op de markt brengen: het kaakje. Als proef probeert hij drie smaakvarianten die hij I, II en III nummert. Hij vraagt honderd mensen om te komen proeven. In het venndiagram zie je de verdeling van hun oordeel. Het venndiagram is nog niet af.

Bereken de kans dat een proever één van de drie smaakvarianten positief beoordeelt, maar de andere twee smaakvarianten niet.



Figuur 2.4

Opgave 15: Oprichters, oplichters en opzichters

Een bestuur van 25 personen bestaat uit oprichters, oplichters en opzichters. Sommige leden hebben meer dan één van die kwaliteiten. Er zijn 10 oprichters, 11 oplichters en 15 opzichters. 1 persoon is zowel oprichter als oplichter en opzichter. 3 zijn oprichter en oplichter (en misschien ook opzichter) en 4 zijn oprichter en opzichter (en misschien oplichter).

a Maak op grond van deze gegevens een Venndiagram.

b Hoe groot is de kans dat een willekeurig bestuurslid keurig is (geen oplichter)?

c Hoe groot is de kans dat een willekeurig bestuurslid oprichter is? Dat hij oplichter is? Dat hij beide is?

d Bepaal de kans dat een willekeurig bestuurslid oprichter of oplichter is.

e De kans dat een bestuurslid oprichter, oplichter of opzichter is, is natuurlijk 1. Iemand zegt: “Die kans moet de kans zijn dat hij oprichter of oplichter is, plus de kans dat hij opzichter is.” Redeneren helpt niet, dus toon hem dat zijn resultaat niet goed kan zijn en vertel hem dan hoe het wel moet.

Testen

Opgave 16

In een vaas zitten 9 balletjes, 3 rode, 3 blauwe en 3 gele. Ze zijn ook genummerd, van elke kleur draagt één balletje nummer 1, één balletje nummer 2 en één balletje nummer 3. Er wordt aselect één balletje getrokken. Bepaal de kans dat:

a het balletje niet rood is;

b het balletje rood is of nummer 2 heeft;

c het balletje niet blauw is of niet nummer 3 heeft.

Opgave 17

Van de leerlingen van een groep staat 70% voldoende voor wiskunde, 63% staat voldoende voor natuurkunde en 43% staat voldoende voor beide vakken.

a Hoeveel procent staat voldoende voor minstens een van beide vakken?

b Hoeveel procent staat onvoldoende voor beide vakken?

c Hoeveel procent staat voldoende voor wiskunde en onvoldoende voor natuurkunde?

d Hoeveel procent staat voldoende voor wiskunde of onvoldoende voor natuurkunde?

1.3 Kansen vermenigvuldigen

Inleiding

Je hebt kennis gemaakt met regels voor de kansrekening. Vooral als het gaat om grotere hoeveelheden (52 verschillende speelkaarten) en meerdere trekkingen zijn dergelijke regels nuttig.
Daarom leer je nu wanneer je kansen mag vermenigvuldigen. Hierbij is het vaasmodel weer erg handig.

Figuur 3.1

Je leert in dit onderwerp

- de regels voor het vermenigvuldigen van kansen;
- werken met afhankelijke en onafhankelijke gebeurtenissen;
- voorwaardelijke kansen berekenen.

Voorkennis

- kansen bepalen met behulp van een kansboom;
- het vaasmodel met of zonder teruglegging voor het berekenen van kansen.

Verkennen

Opgave V1

Een volledig kaartspel kent 52 kaarten, van elke ‘kleur’ evenveel.
Je trekt aselect uit zo'n kaartspel wee kaarten.

a Bereken de kans op twee hartenkaarten.

b Bereken de kans op een hartenkaart en een aas.

Uitleg 1

Je trekt aselect twee kaarten uit een volledig kaartspel. Je kunt dit opvatten als tegelijk twee kaarten trekken of als twee kaarten na elkaar trekken, maar zonder terugleggen. Dit betekent dat de trekking van de tweede kaart ‘afhankelijk’ is van de trekking van de eerste kaart: bij de tweede trekking is er een kaart minder om uit te kiezen. De kans op een aas bij de eerste kaart is $\frac{4}{52}$.
Wanneer de eerste kaart inderdaad een aas was is de kans op een aas bij de tweede kaart is $\frac{3}{51}$.

Hierbij past een kansboom met twee lagen.

De kans op twee azen is $\text{P}(2 \text{ azen}) = \frac{4}{52} \cdot \frac{3}{51} = \frac{12}{572}$.

Je ziet dat je niet zonder meer de kans op twee azen kunt berekenen door de kans op één aas met zichzelf te vermenigvuldigen.

Figuur 3.2

Dat gaat wel als het om trekking met terugleggen gaat, maar niet als het om trekking zonder terugleggen gaat. Bij trekking zonder terugleggen veranderen de kansen na trekking.

Voor de kans op twee azen bij trekking zonder terugleggen geldt:
$P(2 \text{ azen}) = P(\text{aas}) \cdot P(\text{aas} \mid \text{eerste keer aas})$.
De verticale streep | betekent hier: 'gegeven dat' of 'als wat hierna staat al gebeurd is'.
$P(\text{aas} \mid \text{eerste keer aas})$ is een 'voorwaardelijke kans'.
Bij het berekenen ervan moet je rekening houden met wat er eerder is gebeurd.

Opgave 1

Uit een compleet spel speelkaarten worden aselect en zonder terugleggen twee kaarten getrokken.

a Waarom is de tweede trekking van de tweede kaart afhankelijk van de eerste trekking?

b Hoe groot is de kans dat je de eerste keer een harten- en de tweede keer een schoppenkaart trekt?

c Hoe groot is de kans op een harten- en een schoppenkaart?

d Wat is een voorwaardelijke kans?

e Bereken de volgende voorwaardelijke kans:
P(tweede trekking is een hartenkaart | eerste trekking een schoppenkaart).

f Hoe groot is de kans op een vrouw en een heer?

Uitleg 2

Als je gegevens hebt met twee verschillende variabelen die je in één overzicht wilt verzamelen, maak je een 'kruistabel'.

Stel je hebt de gegevens van vijfhonderd Nederlandse vakantiegangers die samen een reis naar Turkije maakten. Je wilt de vakantiegangers indelen naar de variabele *geslacht* en naar de variabele *leeftijdscategorie*, namelijk 'jong', 'middelbaar' en 'oud'. Het resultaat is weergegeven in de kruistabel. Nu zie je meteen dat er 107 mannen van middelbare leeftijd deelnamen aan de Turkije-reis.

	jong	middelbaar	oud	totaal
man	108	107	24	239
vrouw	103	130	28	261
totaal	211	237	52	500

Tabel 3.1

Opgave 2

Op een school is onderzocht hoeveel leerlingen er roken. In de tabel vind je de resultaten van dat onderzoek.

a Hoeveel vrouwelijke leerlingen roken op deze school?

b Bepaal de kans dat een willekeurig meisje van deze school rookt.

c Bepaal de kans dat een willekeurige rokende leerling van deze school een meisje is.

rookgedrag	mannen	vrouwen	totaal
roken	105	134	239
niet roken	475	486	961
totaal	580	620	1200

Tabel 3.2

Opgave 3

Voorwaardelijke kansen komen geregeld voor als je kansen berekent bij frequenties in kruistabellen. Een voorbeeld is een onderzoek naar de Mantoux-test middels een steekproef onder een grote groep personen. De Mantoux-test is een huidtest die wordt gebruikt om na te gaan of iemand tuberculose heeft. Vrijwel alle personen die aan tuberculose lijden, laten een reactie op deze huidtest zien. Maar ook een zeer klein deel van de personen die niet aan tuberculose lijdt, vertoont die reactie.

De kruistabel laat dat zien.

Mantoux-test	tuberculose	geen tuberculose	totaal
reactie	98	99	197
geen reactie	2	9801	9803
totaal	100	9900	10000

Tabel 3.3

a Bereken in vier decimalen nauwkeurig de kans dat iemand die een reactie vertoont op de Mantoux-test, ook inderdaad aan tuberculose lijdt.

b Bereken in vier decimalen nauwkeurig de kans dat iemand die geen reactie vertoont, toch aan tuberculose lijdt.

Theorie en voorbeelden

Om te onthouden

Twee gebeurtenissen zijn **onafhankelijk** als de kans dat de ene gebeurtenis plaatsvindt, niet verandert als de andere gebeurtenis plaatsvindt, en andersom.
Is G_1 een gebeurtenis bij een kansexperiment en G_2 een gebeurtenis bij een tweede kansexperiment en zijn de kansen van het tweede experiment **onafhankelijk** van de uitkomst van het eerste, dan is:

$\text{P}(G_1 \text{ en } G_2) = \text{P}(G_1) \cdot \text{P}(G_2)$

In het vaasmodel is dit het geval als je **met terugleggen** meerdere balletjes trekt.

Is G_1 een gebeurtenis bij een kansexperiment en G_2 een gebeurtenis bij een tweede kansexperiment en zijn de kansen van het tweede experiment **afhankelijk** van de uitkomst van het eerste, dan is:

$\text{P}(G_1 \text{ en } G_2) = \text{P}(G_1) \cdot \text{P}(G_2|G_1)$

In dit geval is $\text{P}(G_2|G_1)$ een **voorwaardelijke kans**, namelijk de kans op G_2 onder de voorwaarde dat G_1 eerst heeft plaatsgevonden.
In het vaasmodel is dit het geval als je **zonder terugleggen** meerdere balletjes trekt.

De regel $\mathrm{P}(G_1 \text{ en } G_2) = \mathrm{P}(G_1) \cdot \mathrm{P}(G_2|G_1)$ heet de **algemene productregel voor kansen** omdat hij ook geldig is voor onafhankelijke gebeurtenissen.
Dan is namelijk $\mathrm{P}(G_2|G_1) = \mathrm{P}(G_2)$.

Als je gegevens hebt met twee verschillende variabelen die je in één overzicht wilt verzamelen, maak je een **kruistabel**.

Voorbeeld 1

Uit een vaas met zes rode en vier blauwe knikkers worden met teruglegging twee knikkers getrokken.
Hoe groot is de kans op een rode en een blauwe knikker?

Antwoord

Het maakt bij de tweede trekking niet uit of de knikker die als eerste getrokken is, rood of blauw was. Door het terugleggen is immers de oorspronkelijke situatie weer hersteld. De tweede trekking is daarom onafhankelijk van de eerste.

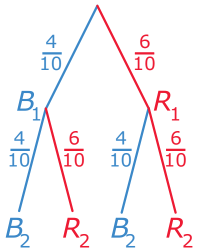
Je kunt dus de productregel voor onafhankelijke kansen gebruiken:

- De kans dat je eerst een rode en dan een blauwe knikker trekt, is:

  $\mathrm{P}(R_1 \text{ en } B_2) = \mathrm{P}(R_1) \cdot \mathrm{P}(B_2) = \frac{6}{10} \cdot \frac{4}{10} = \frac{24}{100}$
- De kans dat je eerst een blauwe en dan een rode knikker trekt, is:

  $\mathrm{P}(B_1 \text{ en } R_2) = \mathrm{P}(B_1) \cdot \mathrm{P}(R_2) = \frac{4}{10} \cdot \frac{6}{10} = \frac{24}{100}$

Beide gebeurtenissen sluiten elkaar uit: $\mathrm{P}(R \text{ en } B) = \frac{24}{100} + \frac{24}{100} = \frac{48}{100}$.



Figuur 3.3

Opgave 4

Bekijk **Voorbeeld 1**.

a Bereken de kans op twee blauwe knikkers.

b Bereken de kans op twee rode knikkers.

c Waarom is hier sprake van onvoorwaardelijke kansen?

Voorbeeld 2

Uit een vaas met zes rode en vier blauwe knikkers worden zonder teruglegging twee knikkers getrokken.
Hoe groot is de kans op een rode en een blauwe knikker?

Antwoord

Het maakt bij de tweede trekking verschil of de knikker die als eerste getrokken is, rood of blauw was. Door de knikkers niet terug te leggen, is immers de oorspronkelijke situatie gewijzigd. De tweede trekking is daarom afhankelijk van de eerste.

Je kunt dus de productregel voor afhankelijke kansen gebruiken:

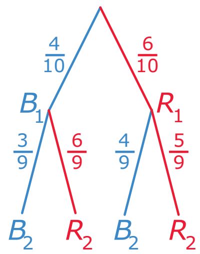
- De kans dat je eerst een rode en dan een blauwe knikker trekt, is:

  $P(R_1 \text{ en } B_2) = P(R_1) \cdot P(B_2|R_1) = \frac{6}{10} \cdot \frac{4}{9} = \frac{24}{90}$

- De kans dat je eerst een blauwe en dan een rode knikker trekt, is:

  $P(B_1 \text{ en } R_2) = P(B_1) \cdot P(R_2|B_1) = \frac{4}{10} \cdot \frac{6}{9} = \frac{24}{90}$

Beide gebeurtenissen sluiten elkaar uit: dus $P(R \text{ en } B) = \frac{24}{90} + \frac{24}{90} = \frac{48}{90}$.



Figuur 3.4

Opgave 5

Bekijk **Voorbeeld 2**.

a Bereken de kans op twee blauwe knikkers.

b Bereken de kans op twee rode knikkers.

c Bereken de voorwaardelijke kans $P(B_2|B_1)$.

d Waarom is het berekenen van $P(B_2|R_2)$ zinloos?

Voorbeeld 3

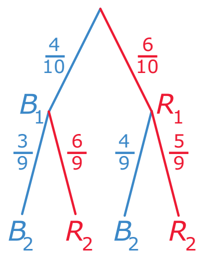
Uit een vaas met zes rode en vier blauwe knikkers worden zonder teruglegging twee knikkers getrokken.
Hoe groot is de kans dat de tweede knikker rood is?

Antwoord

Het maakt bij de tweede trekking verschil of de eerst getrokken knikker rood of blauw was. Door niet terug te leggen, is immers de oorspronkelijke situatie gewijzigd. De tweede trekking is daarom afhankelijk van de eerste.

De gevraagde kans is:

$P(R_2) = P(R_1 \text{ en } R_2) + P(B_1 \text{ en } R_2) =$
$= P(R_1) \cdot P(R_2|R_1) + P(B_1) \cdot P(R_2|B_1) =$
$= \frac{6}{10} \cdot \frac{5}{9} + \frac{4}{10} \cdot \frac{6}{9} = \frac{54}{90}$



Figuur 3.5

Opgave 6

In **Voorbeeld 3** gaat het om de tweede knikker.

Hoe groot is de kans dat de tweede knikker blauw is?

Opgave 7

Je hebt een vaas met zeven rode, vijf witte en acht blauwe knikkers. Je trekt hieruit in één greep drie knikkers.

a Bereken de kans op drie rode knikkers.

b Bereken P (derde knikker blauw | eerste twee knikkers rood).

c Bereken P (derde knikker blauw en eerste twee knikkers rood).

d Bereken de kans op twee rode knikkers.

e Bereken de kans op drie knikkers van verschillende kleur.

Verwerken

Opgave 8

Uit een compleet spel speelkaarten worden aselect twee kaarten getrokken zonder terugleggen. In **Uitleg 1** heb je gezien hoe een complete set speelkaarten is opgebouwd.

a Hoe groot is de kans op twee schoppenkaarten?

b Hoe groot is de kans dat de tweede kaart een schoppenkaart is?

Opgave 9

In West-Europa heeft 40% van de bevolking bloedgroep A, 10% bloedgroep B, 5% bloedgroep AB en 45% bloedgroep O. Voor de Rhesus-factor geldt: 85% is Rh-positief en 15% is Rh-negatief, ongeacht de bloedgroep waartoe men behoort.

a Bereken het percentage West-Europeanen dat bloedgroep A heeft en Rh-positief is.

b Bereken het percentage West-Europeanen dat bloedgroep O heeft en Rh-negatief is.

c Bereken het percentage West-Europeanen dat Rh-negatief is en niet bloedgroep O heeft.

d Welke van de acht combinaties van bloedgroep en Rh-factor is het zeldzaamst?

Opgave 10

Bij een wandeltocht over vochtig terrein zijn je sokken nat geworden. Onder in je rugzak heb je, los door elkaar, acht droge sokken van vier verschillende paren. Je trekt er één sok uit, en dan steeds weer een tot je de bijpassende sok hebt gekregen. Het is verstandig dat je niet teruglegt.

a Hoe groot is de kans dat je precies bij de vierde sok de bijpassende trekt? Let op dat je nog zeven sokken hebt, want de eerste heb je al gepakt.

b Hoe groot is de kans dat de tweede of de derde sok al de goede is?

Opgave 11

In een doos zitten tien kaarten, elk met een cijfer erop. Er is één kaart met een 1, er zijn twee kaarten met een 2, drie kaarten met een 3 en op vier kaarten staat een 4. Je trekt zonder terugleggen vier kaarten en legt die van links naar rechts naast elkaar. Je ziet dan een getal van vier cijfers.

a Hoe groot is de kans dat dit getal 1234 is?

b Hoe groot is de kans dat dit getal 4321 is?

c Hoe groot is de kans dat dit getal 3344 is?

d Bij a en b krijg je hetzelfde antwoord. Licht toe waarom elk van de getallen die je met de cijfers 1, 2, 3 en 4 schrijft dezelfde kans heeft.

Opgave 12

In een doos zitten tien kaarten, elk met een cijfer erop. Er is één kaart met een 1, er zijn twee kaarten met een 2, drie kaarten met een 3 en op vier kaarten staat een 4. Je trekt zonder terugleggen vier kaarten en legt die van links naar rechts naast elkaar. Je ziet dan een getal van vier cijfers.

a Laat E het getal zijn dat door de eerste twee cijfers wordt voorgesteld, T het getal dat door de laatste twee cijfers wordt voorgesteld.

Bereken $\text{P}(T = 34|E = 12)$ en $\text{P}(T = 12|E = 34)$.

b Eén kaart is stiekem door iemand gemerkt. Hoe groot is de kans dat die kaart op uiterst links terechtkomt?

c Voor de kans dat de gemerkte kaart als derde wordt getrokken kun je beter niet rekenen, maar nadenken. Resultaat?

d Test de productregel door na te gaan of je daarmee hetzelfde resultaat krijgt.

e Bereken de kans in de volgende drie gevallen:

- het getal begint met een 3;
- het getal eindigt met een 3;
- het getal begint met een 3 en eindigt met een 3.

Opgave 13

De kans op ten minste één zes bij vier keer gooien met een dobbelsteen is groter dan 50%.

a Laat zien dat dit inderdaad zo is.

Chevalier de Méré dacht daarom (in 1654) dat hij ook meer dan 50% kans had op dubbel zes als hij $6 \cdot 4 = 24$ keer met twee dobbelstenen gooide, maar hij kwam bedrogen uit. Zijn vriend Pascal moest hem uit de droom helpen.

b Bereken de kans op minstens één keer dubbel zes als je 24 keer met twee dobbelstenen gooit in procenten. Geef het antwoord in twee decimalen nauwkeurig.

c Hoe vaak moet je minstens gooien met twee dobbelstenen zodat de kans op dubbel zes groter is dan 50%?

Toepassen

Opgave 14: Welk medicijn?

Bij een bepaalde ziekte kunnen twee verschillende medicijnen worden voorgeschreven: medicijn A of medicijn B. In principe wordt altijd (het beste) medicijn A voorgeschreven, maar 10% van de patiënten reageert daar allergisch op en krijgt dan medicijn B. Medicijn A zorgt in 95% van de gevallen voor genezing, medicijn B in 75% van de gevallen.

Een ziekenhuis heeft deze medicijnen aan negenhonderd personen voorgeschreven. Hoe groot is de kans dat een genezen persoon uit deze groep medicijn B heeft gekregen? Geef het antwoord in vier decimalen nauwkeurig.

Opgave 15: Pepmiddel

Je ziet een vraag uit één van de Nationale Wetenschapsquizzen.

> Met een steekproef worden de deelnemers aan de tiende Nationale Wetenschapsquiz op een verboden pepmiddel getest. Stel dat tien procent van de deelnemers het pepmiddel gebruikt. De test is slechts voor negentig procent zuiver.

Een deelnemer blijkt pep-positief.
Hoe groot is de kans dat hij het pepmiddel daadwerkelijk heeft gebruikt?

Voor alle duidelijkheid: met de zinsnede 'de test is slechts voor negentig procent zuiver' wordt bedoeld dat de test in 10% van de gevallen een verkeerde uitslag geeft. Dus 10% van de gebruikers wordt pep-negatief getest en 10% van de niet-gebruikers wordt pep-positief getest.

Testen

Opgave 16

60% van de artikelen die een fabriek maakt is van soort A, 40% is van soort B. Er gaat wel eens iets mis. Van de artikelen van soort A moet 1% worden afgekeurd. Voor soort B is dat zelfs 2%.

a Wat is de kans dat een aselect gekozen artikel van soort A is en wordt afgekeurd?

b Wat is de kans dat een aselect gekozen artikel van soort B is en wordt goedgekeurd?

c Wat is de kans dat een aselect gekozen artikel wordt afgekeurd?

Opgave 17

Je hebt een vaas met 1200 balletjes, 500 rode, 400 witte en 300 blauwe. Bij elke kleur zijn 200 van de balletjes van hout, de andere zijn van plastic. Het verschil is niet te voelen. Er wordt aselect een balletje uit de vaas gepakt.

a Bepaal P(balletje is wit en van plastic).

b Bepaal P (balletje is rood | balletje is van hout) en P (balletje is van hout | balletje is rood).

c Bepaal P (B is rood of B is van hout).

1.4 Toevalsvariabelen

Inleiding

Bij basketbal wordt per speler het schotpercentage bijgehouden. Als iemand een schotpercentage van 25 heeft, scoort hij bij één op de vier doelpogingen. Je kunt dit percentage daarom opvatten als zijn trefkans bij elke doelpoging.
Maar hoe liggen zijn kansen als hij meerdere doelpogingen doet?

Figuur 4.1

Je leert in dit onderwerp

- een kansverdeling maken;
- kansen berekenen waarbij het gaat om minstens of hoogstens;
- het begrip verwachtingswaarde gebruiken.

Voorkennis

- kansen berekenen met behulp van het vaasmodel en een kansboom.

Verkennen

Opgave V1

Bij basketbal wordt per speler het schotpercentage bijgehouden. Als iemand een schotpercentage van 25 heeft, scoort hij bij één op de vier doelpogingen. Je kunt dit percentage daarom opvatten als zijn trefkans bij elke doelpoging. X stelt het aantal scores voor bij drie doelpogingen.

a Welke waarden kan X aannemen?

b Bereken bij elke waarde van X de bijbehorende kans.

c Hoeveel verwacht je van hem?

Uitleg

Bij basketbal wordt per speler het schotpercentage bijgehouden. Als iemand een schotpercentage van 25 heeft, scoort hij bij één op de vier doelpogingen. Je kunt dit percentage daarom opvatten als zijn trefkans bij elke doelpoging.

Om zijn kansen te bepalen bij bijvoorbeeld drie doelpogingen maak je een kansboom.
Als X het aantal treffers bij deze drie doelpogingen voorstelt, kan X de waarden 0, 1, 2, 3 aannemen.
De bijbehorende kansen kun je berekenen vanuit een kansboom.
Bijvoorbeeld:
$\text{P}(X = 2) = 3 \cdot 0{,}25 \cdot 0{,}25 \cdot 0{,}75 \approx 0{,}14$

Zet je al die kansen op een rij, bijvoorbeeld in een tabel, dan is dat een kansverdeling van X.

x	0	1	2	3
$\text{P}(X = x)$	0,42	0,42	0,14	0,02

Tabel 4.1

Gemiddeld heeft hij bij drie doelpogingen:
$0 \cdot 0{,}42 + 1 \cdot 0{,}42 + 2 \cdot 0{,}14 + 3 \cdot 0{,}02 = 0{,}76$ treffers.
Dat is de verwachte score ofwel de verwachtingswaarde bij drie doelpogingen.

Opgave 1

Lees de **Uitleg**. Neem nu aan dat de basketballer vier doelpogingen doet. Zijn schotpercentage blijft 25.

a Stel een kansverdeling op voor het aantal treffers. Benader de kansen in vier decimalen nauwkeurig.

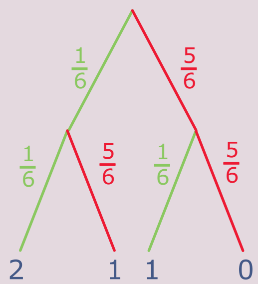
b Bereken het verwachte aantal treffers bij vier doelpogingen.

Theorie en voorbeelden

Om te onthouden

Als een bepaalde variabele X van het toeval afhangt, noem je X een **toevalsvariabele**. Bij elke waarde die X kan aannemen, kun je de bijbehorende kans berekenen (vanuit een kansboom).
Zet je al die kansen op een rij, bijvoorbeeld in een tabel, dan is dat een **kansverdeling** van X.

Noem je bij het werpen met twee dobbelstenen het aantal zessen X, dan is X een voorbeeld van zo'n toevalsvariabele. De bijbehorende kansverdeling haal je uit de kansboom.



Figuur 4.2

x	0	1	2
$\text{P}(X = x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

Tabel 4.2

Ga na dat de som van alle kansen in zo'n kansverdeling 1 is.

Gemiddeld komt er per worp met twee dobbelstenen
$0 \cdot \frac{25}{36} + 1 \cdot \frac{10}{36} + 2 \cdot \frac{1}{36} = \frac{12}{36} = \frac{1}{3}$ keer een zes voor.
Dat heet de **verwachtingswaarde** van het aantal zessen bij het werpen met twee dobbelstenen: bij gemiddeld één op elke drie worpen (met twee dobbelstenen) komt een zes voor. Als je de dobbelsteen maar vaak genoeg werpt.

Voorbeeld 1

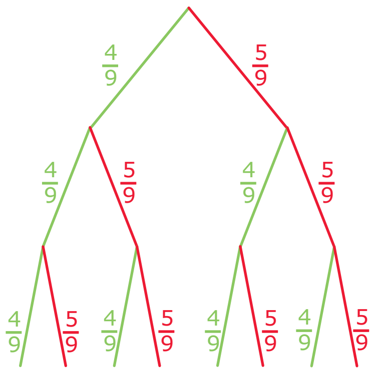
In een groep van vier mannen en vijf vrouwen worden door loten drie taken (wassen, afwassen en auto wassen) verdeeld.
Als één persoon meerdere van die drie taken mag doen, hoe groot is dan de kans dat er hoogstens twee taken door een man worden uitgevoerd?

Antwoord

Je ziet een bijpassende kansboom. Groen betekent dat de taak door een man wordt gedaan, rood dat het een vrouw betreft. Met behulp van de kansboom maak je deze kansverdeling voor het aantal taken M dat door een man wordt uitgevoerd.

m	0	1	2	3
P $(M = m)$	$\frac{125}{729}$	$\frac{300}{729}$	$\frac{240}{729}$	$\frac{64}{729}$

Tabel 4.3



Figuur 4.3

Gevraagd wordt de kans dat er nul, één, of twee taken door een man worden uitgevoerd:

$$\text{P}(M = 0 \text{ of } M = 1 \text{ of } M = 2) = \frac{125}{729} + \frac{300}{729} + \frac{240}{729} = \frac{665}{729}$$

Je kunt ook zo rekenen:

$$\text{P}(M = 0 \text{ of } M = 1 \text{ of } M = 2) = 1 - \text{P}(M = 3) = 1 - \frac{64}{729} = \frac{665}{729}$$

Opgave 2

Bekijk **Voorbeeld 1**.

a Reken de kansen P $(M = 1)$ en P $(M = 2)$ na.

b Bereken de kans dat er minstens twee taken door een man worden uitgevoerd.

c Bereken de kans dat er minder dan twee taken door een man worden uitgevoerd.

Opgave 3

Je gooit met drie dobbelstenen.

a Stel een kansverdeling op voor het aantal keren dat je zes gooit.

b Bereken het verwachte aantal keer dat je zes gooit.

c Beschrijf welke betekenis de verwachtingswaarde heeft.

Voorbeeld 2

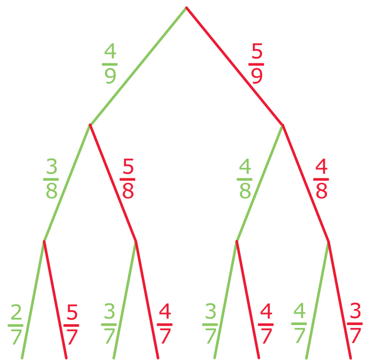
In een groep van vier mannen en vijf vrouwen worden door loten drie taken (wassen, afwassen en auto wassen) verdeeld.
Als elk van die drie taken door een andere persoon moet worden gedaan, hoe groot is dan de kans dat er minstens twee taken door een man worden uitgevoerd? Hoeveel taken worden er naar verwachting door een man uitgevoerd?

Antwoord

Je ziet een bijpassende kansboom. Groen betekent dat de taak door een man wordt gedaan, rood dat het een vrouw betreft. Met behulp van de kansboom maak je deze kansverdeling voor het aantal taken M dat door een man wordt uitgevoerd.

m	0	1	2	3
$\text{P}(M = m)$	$\frac{60}{504}$	$\frac{240}{504}$	$\frac{180}{504}$	$\frac{24}{504}$

Tabel 4.4



Figuur 4.4

Gevraagd wordt de kans dat er twee of drie taken door een man worden uitgevoerd:

$\text{P}(M = 2 \text{ of } M = 3) = \frac{180}{504} + \frac{24}{504} = \frac{204}{504}$

Naar verwachting worden er $0 \cdot \frac{60}{504} + 1 \cdot \frac{240}{504} + 2 \cdot \frac{180}{504} + 3 \cdot \frac{24}{504} = 1\frac{1}{3}$ taken door een man gedaan.

Opgave 4

In **Voorbeeld 2** wordt de verwachtingswaarde voor M uitgerekend.

a Bereken met behulp van een kansverdeling de verwachtingswaarde van het aantal taken dat door een vrouw wordt uitgevoerd.

b Verrast het antwoord bij a je?

Voorbeeld 3

Voor een bepaalde instaptoets worden alleen gehele cijfers van 1 tot en met 10 gegeven. Daarvan zijn jarenlang de resultaten bijgehouden.

cijfer	1	2	3	4	5	6	7	8	9	10
relatieve frequentie (%)	0	0	4	12	20	26	18	11	7	2

Tabel 4.5

Ook dit jaar doet een aantal scholieren deze instaptoets. Welk gemiddeld cijfer verwacht je? En hoeveel procent zal een voldoende halen?

Antwoord

De tabel is op te vatten als een kansverdeling.
De bijbehorende verwachtingswaarde is:
$0 \cdot 1 + 0 \cdot 2 + 0{,}04 \cdot 3 + 0{,}12 \cdot 4 + 0{,}2 \cdot 5 + 0{,}26 \cdot 6 + 0{,}18 \cdot 7 + 0{,}11 \cdot 8 + 0{,}07 \cdot 9 + 0{,}02 \cdot 10 = 6{,}13$
Dus men verwacht een gemiddeld cijfer van 6,1.

Voor het halen van een voldoende moet je minstens een 6 scoren. De kans daarop is: $0{,}26 + 0{,}18 + 0{,}11 + 0{,}07 + 0{,}02 = 0{,}64$.

Opgave 5

In **Voorbeeld 3** wordt een verwachtingswaarde uitgerekend bij statistische gegevens. De plant 'Indigofera australis' plant zich voort door middel van zaden die in zaaddozen aan de plant groeien. Het aantal zaden per zaaddoos kan nogal variëren. Een Britse onderzoeker heeft van een flink aantal zaaddozen het aantal zaden geteld en de resultaten vastgelegd in een tabel.

aantal zaden	3	4	5	6	7	8	9	10	11
frequentie	1	2	8	13	22	45	63	23	1

Tabel 4.6

Neem aan dat deze tabel representatief is voor het aantal zaden per zaaddoos voor deze plant. De relatieve frequenties zijn dan op te vatten als kansen van een toevalsvariabele Z die het aantal zaden per zaaddoos van 'Indigofera australis' voorstelt.

a Maak de bijbehorende tabel met relatieve frequenties in vier decimalen nauwkeurig.

b Bereken $\text{P}(Z > 8)$. Hoeveel procent van de zaaddozen van deze plant heeft meer dan acht zaden? Geef het antwoord in procenten nauwkeurig.

c Hoe groot is de kans op hoogstens vier zaden in een willekeurig gekozen zaaddoos van deze plant? Geef het antwoord in vier decimalen nauwkeurig.

d Hoeveel zaden (in twee decimalen nauwkeurig) verwacht je per zaaddoos?

Verwerken

Opgave 6

Iemand werpt tweemaal een munt op.

Toevalsvariabele K is het aantal keren dat zij kop gooit.

a Maak de tabel met de kansverdeling van K.

b Hoe groot is de kans dat zij hoogstens één keer kop gooit?

Opgave 7

In deze tabel zie je de kansverdeling van toevalsvariabele X:

x	0	1	2	3
$\text{P}(X = x)$	$0{,}15$	$0{,}09$	$0{,}34$	$0{,}42$

Tabel 4.7

Hoe groot is de verwachtingswaarde van X? Geef het antwoord in twee decimalen nauwkeurig.

Opgave 8

Het aantal jongens in een gezin met vier kinderen is een toevalsvariabele J. Ga ervan uit dat de kans op de geboorte van een meisje hetzelfde is als de kans op de geboorte van een jongen.

a Maak een tabel met de kansverdeling van J.

b Hoe groot is de kans dat van de vier kinderen uit een gezin er minstens twee een jongen zijn?

c Hoeveel jongens verwacht je in zo'n gezin?

d Als je 150 van die gezinnen bekijkt, hoeveel jongens verwacht je daar dan tegen te komen? Licht je antwoord toe.

Opgave 9

Uit een vaas met dertig rode en drie groene balletjes wordt vier keer een balletje getrokken.

a X stelt het aantal getrokken groene balletjes voor als telkens wordt teruggelegd. Stel de kansverdeling van X op. Geef de antwoorden in vier decimalen nauwkeurig.

b Y stelt het aantal getrokken groene balletjes voor als niet wordt teruggelegd. Stel de kansverdeling van Y op. Geef de antwoorden in vier decimalen nauwkeurig.

Opgave 10

Uit een klas met zestien meisjes en twaalf jongens worden vier verschillende leerlingen gekozen. M is het aantal meisjes in die groep van vier.

a Welke waarden kan M aannemen?

b Maak een staafdiagram van de kansverdeling van M.

c Bepaal het verwachte aantal meisjes in de groep van vier. Geef het antwoord in één decimaal nauwkeurig.

Opgave 11

Je gooit met twee dobbelstenen en vermenigvuldigt het aantal ogen op de ene dobbelsteen met het aantal ogen op de andere. Dat is de waarde van de toevalsvariabele Z.

a Stel de kansverdeling van Z op.

b Je krijgt de waarde die Z aanneemt, uitbetaald in euro's. Zou je voor dat spel € 12,00 willen inzetten? Hoe groot is dan de kans dat je met één spel winst maakt?

Opgave 12

In een casino wordt maximaal zes keer met een zuivere munt geworpen totdat kruis boven komt. De speler krijgt respectievelijk € 1,00; € 2,00; € 4,00; € 8,00; € 16,00 of € 32,00 uitbetaald als bij worp nummer 1, 2, 3, 4, 5 respectievelijk 6 kruis wordt gegooid. Bij zes keer munt krijgt hij niets.

Welke inzet moet het casino voor dit spel vragen om op de lange duur gemiddeld € 0,50 per spel winst te krijgen?

Toepassen

Opgave 13: Wimbledon

In de finale herenenkel van het tennistoernooi van Wimbledon wordt gespeeld om 'best of five': wie het eerst drie sets heeft gewonnen is kampioen. Na hoogstens vijf sets is er dus een winnaar; het kan al na drie sets. Neem je aan dat beide finalisten even sterk zijn en kans 50% hebben om een set te winnen, dan is het aantal in de finale gespeelde sets een toevalsvariabele S.

a Maak daarvan een kansverdeling en bereken het verwachte aantal sets.

b Neem aan dat het toernooi van Wimbledon al honderd keer is gespeeld. Hoeveel sets zijn er dan naar verwachting in totaal in de finales gespeeld?

De werkelijke gegevens leren toch anders, zie de tabel over 90 finales.

partijlengte	3 sets	4 sets	5 sets
aantal keer	44	22	24

Tabel 4.8

c Bepaal de experimentele kansverdeling en verwachtingswaarde van S.

De oorspronkelijke aanname was dus niet zo goed. Stel je nu voor dat de kans om de eerste set te winnen 50% blijft, maar de kans om de set na een gewonnen set te winnen 70% is (de 'winning mood').

d Maak nu opnieuw een kansverdeling (bekijk zorgvuldig alle gevallen).

e Bereken het verwachte aantal sets bij de nieuwe kansverdeling.

Opgave 14: Vogelsoorten

Vogeldeskundigen willen weten welke vogelsoorten in een bepaald gebied leven. Een eenvoudige manier om daarachter te komen, is het maken van een ronde door dat gebied en alle waargenomen vogels te registreren. Men spreekt van een registratie-effectiviteit van 100% wanneer alle aanwezige vogels opgemerkt worden. In de praktijk blijkt de registratie-effectiviteit per ronde slechts 60% te zijn, de overige 40% van de totale vogelpopulatie wordt niet opgemerkt. De Zweedse vogeldeskundige Anders Enemar stelt dat de registratie-effectiviteit door het maken van drie ronden zodanig wordt verhoogd, dat men vrijwel zeker mag aannemen dat alle vogelsoorten zijn opgemerkt. Hij neemt daarbij aan dat iedere aanwezige vogel bij elke ronde 60% kans heeft om opgemerkt te worden.

a Bereken hoeveel procent van de totale populatie naar verwachting na drie ronden nog niet zal zijn opgemerkt.

Na drie ronden is de vogelpopulatie verdeeld in vier categorieën: I, II, III, IV.

I niet opgemerkt;

II één keer opgemerkt;

III twee keer opgemerkt;

IV drie keer opgemerkt.

b Welke van deze vier categorieën bevat de meeste exemplaren? Licht je antwoord toe met een berekening.

Stel dat er bij iedere ronde ongeveer 450 vogels worden opgemerkt.

c Bereken hoeveel vogels er ongeveer bij de derde ronde voor het eerst worden opgemerkt.

(bron: examen wiskunde A in 1990, eerste tijdvak)

Testen

Opgave 15

Er wordt met 4 munten geworpen. Het aantal keren dat kruis boven komt is een toevalsvariabele K.

a Bepaal de kansverdeling van K.

b Bepaal de verwachtingswaarde van K.

Je herhaalt het werpen met deze vier munten 200 keer.

c Hoeveel keer zul je daarbij $K = 3$ aantreffen?

Opgave 16

Het bestuur van de korfbalclub telt zeven leden, vier vrouwen en drie mannen. Door loten wordt daaruit een dagelijks bestuur van drie leden gekozen. Het aantal vrouwen in het dagelijkse bestuur is een toevalsvariabele Z.

a Maak een kansverdeling van Z.

b Hoe groot is de kans dat er minstens twee vrouwen in het dagelijks bestuur zitten?

c Bereken het verwachte aantal vrouwen in het bestuur.

1.5 Totaalbeeld

Samenvatten

Je hebt nu het onderwerp **Kansrekening** doorgewerkt. Er moet een totaalbeeld van deze leerstof ontstaan...
Ga na, of je al de bij dit onderwerp horende begrippen kent en weet wat je er mee kunt doen. Ga ook na of je de activiteiten die staan genoemd kunt uitvoeren. Maak een eigen samenvatting!

Begrippenlijst

- vaasmodel — kansboom — trekking met/zonder teruglegging
- kansexperiment — uitkomstenverzameling, gebeurtenis — elkaar wederzijds uitsluitende gebeurtenissen
- algemene productregel voor kansen — (on)afhankelijke gebeurtenissen — voorwaardelijke kans
- toevalsvariabele — kansverdeling — verwachtingswaarde

Activiteitenlijst

- kansen berekenen m.b.v. kansbomen — het vaasmodel gebruiken
- kansen berekenen met de optelregel
- kansen berekenen de productregel
- een kansverdeling maken — de verwachtingswaarde berekenen — kansen berekenen waarin het gaat om minstens of hoogstens een bepaald aantal

Achtergronden

In het midden van de zeventiende eeuw is de kansrekening ontstaan in een briefwisseling van de twee grote Franse wiskundigen **Pierre de Fermat (1601—1665)** en **Blaise Pascal (1623—1662)**.

Halverwege de zeventiende eeuw kreeg Pascal het onderstaande kansprobleem voorgelegd door de Franse edelman (en verwoed gokker) Chevalier de Méré.

De Méré speelde in de Franse 'salons' vaak een dobbelspel waarbij de 'bank' won als een speler bij het werpen met één zuivere dobbelsteen bij 4 worpen tenminste één zes gooit. Hij bedacht daarop een variant waarbij de bank wint wanneer bij 24 worpen met twee zuivere dobbelstenen tenminste één keer dubbel-zes voorkwam. De Méré dacht dat er bij beide situaties voor de bank dezelfde kans op winst bestond: in het eerste geval 4/6 en in het tweede geval 24/36 (want bij twee dobbelstenen zijn er 36 mogelijkheden), en dat is beide hetzelfde. In de praktijk bleek dit echter niet op te gaan, de tweede situatie was voor de bank ongunstig. De vraag was hoe dat kwam.

Pascal stortte zich op deze problemen en in een briefwisseling met Pierre de Fermat losten zij ze op. Daarbij ontwikkelden ze de **basisprincipes van de kansrekening**. In feite zijn Pascal en Fermat de grondleggers van de kansrekening zoals wij die tegenwoordig nog steeds beoefenen. Zij werkten echter met kansen in termen van verhoudingen als 1 : 6 en niet (zoals wij tegenwoordig doen) met breuken.

Het eerste echte leerboek over kansrekening is echter geschreven door de Nederlandse geleerde **Christiaan Huygens (1629—1695)**. In zijn 'Van Rekeningh in Spelen van Geluck' introduceerde Huygens het vaasmodel met zwarte en witte schijven, zoals bij het damspel. Allerlei vraagstukken uit de kansrekening herleidde hij tot dat vaasmodel.

Figuur 5.1

Later schreef de Franse wiskundige **Abraham de Moivre (1667—1754)**, die sinds 1685 in Londen woonde en een goede vriend was van Isaac Newton, zijn beroemde 'The Doctrine of Chance', waarin hij de basisbegrippen van de kansrekening voortzette naar een theorie over kansverdelingen.

De moderne opbouw van de kansrekening is van betrekkelijk recente datum. De Russische wiskundige **Andrej Kolmogorov (1903—1987)** zorgde voor een precieze wiskundige theorie.

Testen

Opgave 1

Vertaal de volgende situaties in een vaasmodel en bereken de kans.

a Bij de presidentsverkiezingen in de Verenigde Staten in 2000 ging de verkiezingsstrijd tussen de presidentskandidaten Al Gore en George Bush. Gore had op zeker moment ongeveer veertig procent van de kiezers achter zich en Bush ook. De overige kiesgerechtigde Amerikanen zouden niet gaan stemmen. Je komt vier toeristen uit de Verenigde Staten tegen. Hoe groot is de kans dat ze op dat moment alle vier op Gore zouden stemmen?

b Bij een reddingsoperatie moeten drie vrijwilligers een gebouw in. Er zijn twee brandweerkorpsen uitgerukt: korps A met tien leden en korps B met vijftien leden. Alle leden van de brandweerkorpsen melden zich als vrijwilliger. De drie vrijwilligers worden door het lot aangewezen. Hoe groot is de kans dat ze alle drie bij korps A horen? Geef de antwoorden in vier decimalen nauwkeurig.

c Je gooit met drie gewone dobbelstenen. Hoe groot is de kans op een som van vijftien ogen?

d Je bent je pincode vergeten. Die pincode bestaat uit vier cijfers en alle mogelijkheden zijn toegestaan. Je wilt geld uit de geldautomaat halen. Je toetst zomaar een pincode in. Hoe groot is de kans dat deze de juiste is?

Opgave 2

In een vaas zitten twintig balletjes: tien rode, vijf gele en vijf blauwe. Uit die vaas worden aselect tegelijk drie balletjes gehaald.

a Maak een kansboom bij deze situatie.

b Hoe groot is de kans dat er twee rode en één blauw balletje worden getrokken? Geef het antwoord in vier decimalen nauwkeurig.

c Hoe groot is de kans op één balletje van elke kleur? Geef het antwoord in vier decimalen nauwkeurig.

d Hoe groot is de kans op hoogstens twee rode balletjes? Geef het antwoord in vier decimalen nauwkeurig.

Opgave 3

Uit een groep van tien mannen en zestien vrouwen worden door loten vier vertegenwoordigers aangewezen. M stelt het aantal aangewezen mannen voor, V het aantal aangewezen vrouwen.

a Bereken $\text{P}(M = 0)$ en $\text{P}(M = 1)$ in procenten, en rond af op twee decimalen.

b Maak een kansverdeling voor M in procenten, en rond af op twee decimalen.

c Maak een kansverdeling voor V in procenten, en rond af op twee decimalen.

d Bereken de verwachtingswaarde van M en V. Rond af op twee decimalen.

Opgave 4

In een zeker gebied in Afrika beschikt zestig procent van de bewoners over goed drinkwater. Acht procent van de bewoners heeft een bepaalde darmparasiet; van hen heeft slechts één op de vier goed drinkwater.

a Hoe groot is de kans dat een willekeurige bewoner goed drinkwater en toch die darmparasiet heeft?

b Hoe groot is de kans dat een willekeurige bewoner goed drinkwater en niet die darmparasiet heeft?

c Hoe groot is de kans dat een willekeurige bewoner goed drinkwater of niet die darmparasiet heeft?

d De kans dat een bewoner met goed drinkwater die parasiet heeft, zal kleiner zijn dan de kans dat een bewoner zonder goed drinkwater die parasiet heeft. Hoe groot zijn die kansen in procenten? Geef je antwoorden in twee decimalen nauwkeurig.

Opgave 5

Bij een gezondheidsenquête, uitgevoerd door het Centraal Bureau voor de Statistiek, waren vragen opgenomen over linkshandigheid. Van linkshandige meisjes en jongens in de leeftijd van tien tot twintig jaar is nagegaan of de ouders links- of rechtshandig zijn. Het resultaat hiervan staat in de tabel.

CBS	één van de ouders of beide ouders linkshandig	beide ouders rechtshandig
aantal meisjes linkshandig	32	72
aantal jongens linkshandig	40	96

Tabel 5.1

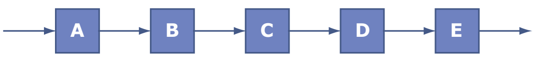
Een linkshandige jongen en een linkshandig meisje (uit die leeftijdscategorie) beginnen een relatie. Na verloop van tijd maken de ouders van beide kinderen kennis met elkaar. Die ouders blijken alle vier rechtshandig te zijn.

Hoe groot is die kans op vier rechtshandige ouders hier? Geef het antwoord in vier decimalen nauwkeurig.

(bron: examen wiskunde A in 1991, eerste tijdvak)

Opgave 6

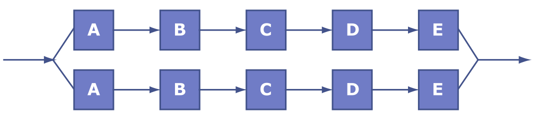
Bij een ingewikkeld apparaat is vaak een keten van onderdelen nodig om het geheel te laten functioneren. Daarbij is de betrouwbaarheid van een keten (zoals in onderstaande figuur) kleiner dan de betrouwbaarheid van de afzonderlijke delen. Dat komt doordat het uitvallen van één onderdeel het uitvallen van de gehele keten tot gevolg heeft. Bekijk een keten van 5 onderdelen (A, B, C, D, E), die elk een kans van 10% hebben om uit te vallen, oftewel elk een betrouwbaarheid hebben van 90%.



Figuur 5.2

a Laat zien dat de betrouwbaarheid van deze keten ongeveer 60% is.

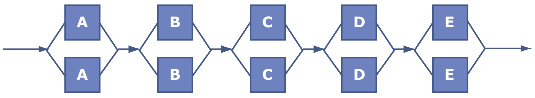
Men kan de betrouwbaarheid vergroten door naast de keten van bovenstaande figuur nog zo'n keten te schakelen (zie de volgende figuur). Dit heeft het voordeel dat als één keten uitvalt het systeem toch blijft functioneren.



Figuur 5.3

b Bereken de betrouwbaarheid van dit systeem.

Een nog beter systeem krijgt men door de 10 onderdelen zo te schakelen als de volgende figuur aangeeft. Elk van de tien onderdelen heeft weer een betrouwbaarheid van 90%.



Figuur 5.4

c Bereken de betrouwbaarheid van dit laatste systeem.

Opgave 7

De verkiezing voor de Tweede Kamer in januari 2017 leverde de volgende zetelverdeling op.
K is een willekeurig kamerlid.

TK 2017	VVD	PVV	CDA	D66	GL	SP	PvdA	CU	PvdD	50+	SGP	Denk	FvD
mannen	23	14	13	12	6	9	4	3	2	2	3	3	2
vrouwen	10	6	6	7	8	5	5	2	3	2	0	0	0
totaal	33	20	19	19	14	14	9	5	5	4	3	3	2

Tabel 5.2

a Bepaal de kans dat K een vrouw is.

b Hoe groot is de kans dat K een vrouw is als K lid is van het CDA? Lid is van de PvdA? Lid is van het CDA of van de PvdA?

c Bepaal door tellen en ook met de productregel de kans dat K een vrouwelijk lid van het CDA is.

d Hoe groot is de kans dat een vrouwelijk kamerlid bij de VVD hoort? Niet bij de VVD hoort? Bij de SGP hoort?

e Welke kans is het grootste, de kans dat een VVD kamerlid een vrouw is of de kans dat een vrouwelijk kamerlid van de VVD is?

Toepassen

Opgave 8: Chuck-a-luck

‘Chuck-a-luck’ is een kansspel waarbij wordt geworpen met drie dobbelstenen. Het wordt in veel casino's gespeeld. Een casino is vooral geïnteresseerd in de verwachtingswaarde, veel spelers denken alleen aan hun kansen (sukkels...).

Bij dit spel kies je een bepaalde uitkomst voor het aantal ogen op één dobbelsteen. Komt jouw aantal bij een worp met drie stenen één keer voor krijg je de inleg terug, komt het twee keer voor krijg je je inleg twee keer terug en komt het drie keer voor dan krijg je je inleg 10 keer terug.

Stel je voor dat W je winstbedrag is. Per ingelegde euro heeft W de waarden -1, 0, 1 en 9. Met behulp van een kansboom maak je daarbij een kansverdeling en bereken je de verwachtingswaarde. Vooral die verwachtingswaarde is interessant voor een casino: het is het gemiddelde bedrag dat ze per ingelegde euro moeten uitbetalen.

Figuur 5.5

a Stel een bij dit spel passende kansverdeling op.

b Bereken de bijbehorende verwachtingswaarde.

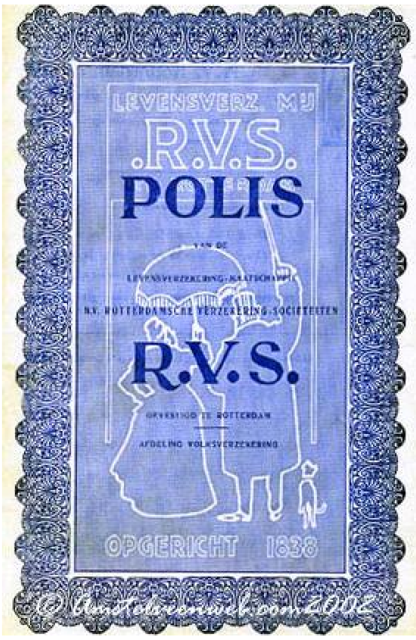
c Wat adviseer je een casino dat dit spel wil invoeren?

Opgave 9: Sterftetabellen

Verzekeringsmaatschappijen gebruiken veel kansrekening. Bij het afsluiten van een levensverzekering willen verzekeraars weten wat de kans is dat de verzekerde binnen een bepaalde tijd overlijdt. Daarbij wordt gebruik gemaakt van tabellen zoals deze **sterftetabel**. Deze tabellen zijn gebaseerd op statistisch onderzoek en worden van tijd tot tijd bijgesteld.

Je ziet in deze tabel bijvoorbeeld dat van elke $10.000.000$ geboren mannen er na 35 jaar nog $9.804.341$ in leven zijn. Na 50 jaar zijn dat er nog $9.545.529$. De kans dat iemand van 35 jaar oud 50 jaar wordt is dan: $\frac{9.545.529}{9.804.341} \cdot 100\% \approx 97{,}4\%$.

Zo kun je ook zijn kans bepalen dat hij 70 jaar wordt, en 71 jaar, etc. En daarmee bereken je zijn levensverwachting en weet de verzekeringsmaatschappij hoeveel jaar er gemiddeld aan iemand van 35 jaar oud nog moet worden uitbetaald vanaf het moment dat zijn levensverzekering tot uitkering komt.



Figuur 5.6

a Hoeveel procent van de mensen die de leeftijd van 25 jaar hebben bereikt, sterven voor hun dertigste?

b Hoeveel procent van de mensen die 60 jaar worden, sterven voor hun zeventigste?

c Gebruik nog eens de gegeven sterftetabel. Bereken in vier decimalen nauwkeurig de kans dat iemand die 50 jaar wordt, ook nog zijn zeventigste verjaardag zal halen.

d Bereken ook de kans dat iemand die 50 jaar wordt, binnen 20 jaar zal sterven.

e Bereken de levensverwachting van een vijftigjarige man met behulp van deze tabel in Excel.

Stel je een levensverzekering voor waarbij je het verzekerde bedrag in één keer krijgt uitbetaald wanneer je op de afgesproken datum nog in leven bent. Je betaalt de premie in één keer op het moment dat je de levensverzekering afsluit. Stel je voor dat je op 30-jarige leeftijd zo'n levensverzekering afsluit met als verzekerd bedrag € 100.000. Dit bedrag wordt uitgekeerd op het moment dat je 65 jaar wordt en nog in leven bent. De premie kan echter lager zijn dan € 100.000. Anders kun je immers beter zelf het geld op de bank zetten!

f Waarom is dat zo?

g Hoe hoog zou die premie moeten zijn op grond van de gegeven tabel? Licht je antwoord toe, houd rekening met overlevingskansen en rentestand.

Opgave 10: Drie deuren probleem

De winnaar van een quiz mag uit drie deuren er een kiezen. De deuren zien er hetzelfde uit, maar achter één ervan zit de hoofdprijs, achter de andere twee zit niks. Na die keuze wijst de spelleider een andere deur aan en zegt (naar waarheid) dat die leeg is. De winnaar mag nu nog zijn keuze veranderen.

a Bereken de winstkans in het geval dat hij niet wisselt.

b Bereken de winstkans in het geval dat hij wisselt.

Examen

Opgave 11: Wijn proeven

Bij het examen voor vinoloog (wijnkenner) moeten de kandidaten wijnen herkennen door te proeven. Uit een artikel komt de volgende tekst.

> De examenkandidaten hebben zich een jaar lang op deze proeverij voorbereid. Het zijn bijna allemaal vaklui: restauranteigenaars, wijnproevers, slijters. De opdracht lijkt simpel: combineer de 12 op papier genoemde wijnen met het juiste glas. Om te slagen wordt genoegen genomen met 9 juiste combinaties. Dat dit in de praktijk een hels karwei is, blijkt wel uit het geringe aantal kandidaten dat succesvol is: gemiddeld zo'n 30 procent.

In deze opgave kijken we naar de kans dat iemand die helemaal geen verstand van wijnen heeft het examen haalt. Omdat hij uitsluitend gokt, noemen we hem een gokker.
Er staan, volgens bovenstaande tekst, 12 glazen met wijn op tafel. Iedere deelnemer krijgt 12 kaartjes met de namen van die wijnen. De opdracht is: leg bij elk glas het goede kaartje. De gokker legt zijn kaartjes dus in willekeurige volgorde bij de verschillende glazen.

a Op hoeveel verschillende manieren kan de gokker de kaartjes neerleggen?

Om het iets gemakkelijker te maken, heeft de examencommissie de 12 wijnen in 4 groepjes van 3 verdeeld. Bij elk groepje liggen 3 kaartjes met de namen van die 3 wijnen. De opdracht van de kandidaat is om bij elk groepje de kaartjes bij het juiste glas te leggen.

b Stel een kansverdeling op van het aantal door de gokker goed neergelegde kaartjes per groepje van 3.

In deze tabel zie je een mogelijk verloop van het examen. De ‘route’ 3 – 1 – 0 – 3 levert in totaal 7 goed geraden wijnen.

	eerste drietal	tweede drietal	derde drietal	vierde drietal
aantal	0	0	0	0
goed neergelegde	1	1	1	1
kaartjes	3	3	3	3

Tabel 5.3

Om te slagen moeten er minstens 9 wijnen goed geraden worden.

c Bereken de kans dat een gokker slaagt.

(bron: voorbeeldexamen vwo wiskunde A1,2 in 2001)

Opgave 12: Vierkeuzevragen

Bij vierkeuzevragen staan bij elke vraag vier mogelijke antwoorden: A, B, C en D. Slechts één daarvan is juist. Een kandidaat kan één van de vier antwoorden kiezen of de vraag onbeantwoord laten. Bij keuze van het juiste antwoord wordt 1 punt toegekend, in alle andere gevallen 0 punten. Als een kandidaat absoluut niet weet welk antwoord juist is en welke antwoorden onjuist zijn, doet hij er daarom verstandig aan om toch een antwoord te kiezen. Dit leidt tot gokgedrag.
Daarom is ook wel eens geopperd om bij een onjuist antwoord strafpunten te geven. Een kandidaat heeft dan twee keuzes: niets invullen levert 0 punten op; wel iets invullen levert 1 punt op bij een juist antwoord en $-0{,}5$ punt ($0{,}5$ strafpunt) bij een onjuist antwoord.

a Bereken de verwachtingswaarde van de score per vraag bij dit strafpuntensysteem als een kandidaat gokt.

We kijken nu naar een andere manier van toetsen met vierkeuzevragen. Hierbij hoeft de kandidaat niet meer één antwoord te kiezen. In plaats daarvan vraagt men de kandidaat achter elk van de vier mogelijke antwoorden A, B, C en D de subjectieve kans op te schrijven.
Een kandidaat die bijvoorbeeld noteert $p_A = 0{,}2$; $p_B = 0{,}8$; $p_C = 0$; $p_D = 0$ geeft daarmee aan dat hij er vrij zeker van is dat B juist is, maar dat A ook nog zou kunnen, en dat C en D volgens hem zeker fout zijn. De opgeschreven getallen p_A, p_B, p_C en p_D mogen natuurlijk niet negatief zijn en moeten bij elkaar opgeteld 1 zijn.
Bij iedere vraag wordt een score berekend die aangeeft ‘hoe dicht je bij het juiste antwoord zit’.
Als bijvoorbeeld C het juiste antwoord is, dan wordt de score berekend met de volgende formule:

$$\text{score} = 1 - \left(p_A^2 + p_B^2 + (1 - p_C)^2 + p_D^2\right).$$

Voor de gevallen waarbij A, B of D het juiste antwoord is, gelden soortgelijke formules. De maximale score is 1 en de minimale score is -1.
Bij een bepaalde vraag is het juiste antwoord B. Een kandidaat die niet helemaal zeker van zijn zaak is, noteert bij deze vraag als subjectieve kansen: $p_A = 0{,}2$; $p_B = 0{,}7$; $p_C = 0$; $p_D = 0{,}1$.

b Bereken de score voor deze kandidaat bij deze vraag.

Stel dat bij een andere vraag C het juiste antwoord is. Een kandidaat haalt bij deze vraag de minimale score.

c Welke subjectieve kansen kan de kandidaat opgeschreven hebben achter de antwoorden A, B, C en D? Vermeld alle mogelijkheden.

Een kandidaat moet een vraag beantwoorden maar heeft geen idee welk antwoord juist is en welke antwoorden onjuist zijn. Er zijn heel veel mogelijkheden voor de kandidaat om die vraag te beantwoorden:

- Mogelijkheid I:
  De kandidaat zou kunnen gokken op een antwoord door daar 1 achter te schrijven (en dus 0 achter de andere antwoorden). Het antwoord waarbij de kandidaat 1 heeft gezet, kan goed zijn. Dan is de score 1. Als het niet goed is, is de score -1. De verwachte score is dan: $\frac{1}{4} \cdot 1 + \frac{3}{4} \cdot -1 = -0{,}5$.
- Mogelijkheid II:
  Hij kan ook op twee antwoorden gokken door achter ieder van die twee antwoorden $\frac{1}{2}$ te schrijven.
- Mogelijkheid III:
  Hij kan ook op drie antwoorden gokken door achter ieder van die drie antwoorden $\frac{1}{3}$ te schrijven.
- Mogelijkheid IV:
  En tenslotte kan hij ook op alle vier de antwoorden gokken door achter alle antwoorden $\frac{1}{4}$ te schrijven. Deze laatste mogelijkheid levert hem een score van $0{,}25$ op.

Er zijn nog veel meer mogelijkheden om de vraag te beantwoorden. We kijken echter alleen naar de bovengenoemde vier mogelijkheden. De score bij mogelijkheid IV is hoger dan de verwachte score bij mogelijkheid I. Mogelijkheid IV is daarmee een 'verstandiger' strategie dan mogelijkheid I.

d Onderzoek welk van de mogelijkheden II, III en IV de meest verstandige strategie is.

We vergelijken de antwoorden van twee personen op een vierkeuzevraag. Tim snapt niets van de vraag en besluit bij ieder antwoord $0{,}25$ in te vullen. Tom weet zeker dat de antwoorden B en D onjuist zijn. Zijn antwoord op deze vraag zal de vorm hebben die je in de tabel ziet. Hierbij is a een getal tussen 0 en 1 (eventueel 0 of 1).

antwoord	subjectieve kans
A	a
B	0
C	$1 - a$
D	0

Tabel 5.4

Stel dat antwoord C juist is. Of Tom bij deze vraag een hogere score haalt dan Tim hangt af van de gekozen waarde van a.

e Bereken voor welke waarden van a geldt dat Tom bij deze vraag een hogere score haalt dan Tim.

(bron: examen wiskunde A1,2 vwo 2004, eerste tijdvak)

2

Statistiek

2.1 Onderzoeken

Inleiding

Statistiek houdt zich bezig met het verzamelen, uitleggen en presenteren van betrouwbare gegevens over groepen (van bijvoorbeeld mensen). Meestal wordt daarbij maar een deel van de groep onderzocht.

Het woord ‘statistiek’ is afkomstig uit het Latijn: ‘statisticum collegium’. Dit betekent: les over staatszaken. Je zou kunnen zeggen: de analyse van staatsgegevens.

Figuur 1.1

Je leert in dit onderwerp

- de begrippen statistisch onderzoek, steekproef, populatie, aselect en representatief;
- onderscheid maken tussen kwalitatieve en kwantitatieve variabelen en tussen discrete en continue kwalitatieve variabelen;
- een vragenlijst opstellen die betrouwbare gegevens oplevert die goed te verwerken zijn;
- toevalsgetallen genereren en er handig gebruik van maken;
- het verschil tussen beschrijvende en verklarende statistiek.

Voorkennis

- de basistechnieken van het werken met de grafische rekenmachine en/of met Excel.

Verkennen

Opgave V1

In en rond de Randstad zijn er vrijwel elke dag files. Er wordt al jaren onderzoek gedaan hoe dit probleem de wereld uit geholpen kan worden. De verzamelnaam voor het onderzoek in waar files zich voordoen en de vele voorgestelde oplossingen is het ‘fileprobleem’.

a Voor welke beroepsgroepen is de ontwikkeling van het fileprobleem van belang?

b Waarom is statistisch onderzoek noodzakelijk om het ontstaan van files te onderzoeken?

c Wie doen in Nederland statistisch onderzoek?

Opgave V2

Bedenk een opzet voor een onderzoek onder een deel van ouders en leerlingen van de school naar hun gebruik van de fiets. Hierbij wil je uiteindelijk uitspraken kunnen doen over het gebruik van de fiets van alle leerlingen van de school en hun ouders.

Uitleg 1

Het percentage mensen dat (vrijwel) nooit vlees eet is minder dan 5%, bij 17 miljoen Nederlanders dus 850.000 mensen. Dit aantal is wel stijgende en gebaseerd op onderzoeken van onder andere het RIVM, het Landbouw Economisch Instituut, Natuur en Milieu, het Voedingscentrum en Milieu Centraal.
Bron: Nederlandse vegetariërsbond, maart 2017

Belangrijke vragen bij zo'n bewering zijn:

- Waar komen deze gegevens vandaan?
- Hoe betrouwbaar is zo'n uitspraak?

Statistiek houdt zich bezig met het verzamelen, uitleggen en presenteren van betrouwbare gegevens over groepen (van bijvoorbeeld mensen), waarbij meestal maar een deel van de groep wordt onderzocht. Je noemt de groep ‘de populatie’ en het deel van de groep ‘een steekproef’. De gegevens die je verzamelt heten ‘data’.

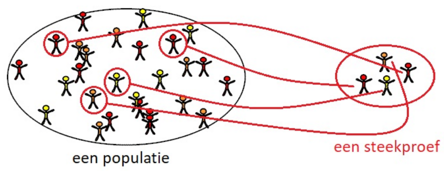
Het kiezen van de steekproefdeelnemers moet ‘aselect’ (willekeurig) plaatsvinden.

De steekproef moet ‘representatief’ zijn: alle soorten deelnemers moeten naar verhouding even vaak in de populatie voorkomen als in de steekproef.

Om te zorgen dat je een goede ‘steekproef’ hebt, moet je ook zorgen dat deze groot genoeg is.

Een statistisch onderzoek begon met de onderzoeksvraag: “Hoeveel procent van de Nederlanders is vegetariër?”

Deze vraag gaat over de variabele ‘aantal Nederlanders dat vegetariër is’. Deze variabele is kwantitatief (in getallen uit te drukken). Daartegenover staan kwalitatieve variabelen zoals geslacht of politieke voorkeur.



Figuur 1.2

Opgave 1

Bestudeer de uitleg. Welke van de volgende steekproeven zijn representatief?

a Om een onderzoek te doen naar het discotheekbezoek onder 14- tot 18-jarigen kies je de leerlingen van jouw eigen klas.

A. representatief
B. twijfelachtig
C. representatief als de steekproef groot genoeg is

b Om de politieke voorkeur van Nederlanders te bepalen, worden aselect uit de telefoongids van Nederland 1250 abonnees getrokken die aan het onderzoek deelnemen.

A. representatief
B. twijfelachtig
C. representatief als de steekproef groot genoeg is

c Om de gemiddelde onderhoudskosten van een bepaald automerk te bepalen, worden de bezitters van zo'n auto via de *Autovisie* opgeroepen om de kosten die zij gemaakt hebben, op te geven.

A. representatief
B. twijfelachtig
C. representatief als de steekproef groot genoeg is

d Om de kwaliteit van diepvrieskippen te bepalen, kopen de onderzoekers 190 diepvrieskippen; van 19 merken steeds 10 aselect getrokken stuks.

A. representatief
B. twijfelachtig
C. representatief als de steekproef groot genoeg is

Opgave 2

In welke gevallen is sprake van een aselecte steekproef?

A. Tien Deventernaren selecteren door uit het bevolkingsregister van Deventer de eerste tien namen die met een H beginnen, te nemen.
B. Een provincie in Nederland kiezen door deze geblinddoekt op een kaart van Nederland aan te wijzen.
C. Vijf havo 4-leerlingen kiezen door uit een zak met gevouwen lootjes met alle achternamen van leerlingen uit 4 havo de eerste vijf te halen.
D. Automobilisten om en om aanhouden en controleren bij de toegangsweg van een dorp vanaf 20:00 tot 22:00 uur.

Opgave 3

Je doet een onderzoek onder jongeren naar hun mening over de smartwatch.

Welke van de hier genoemde onderzoeksmiddelen zijn het meest geschikt?

A. een telefonische enquête
B. een vragenlijst in een meidenblad
C. een vragenlijst via sociale media
D. een vragenlijst op straat vlak bij een winkelcentrum

Uitleg 2

In een dataset kunnen verschillende soorten variabelen zitten. Je kent het onderscheid tussen kwalitatieve en kwantitatieve variabelen al. Met kwantitatieve variabelen kun je rekenen, met kwalitatieve variabelen niet. Met een hobby (kwalitatief) kun je niet rekenen, met een gemiddeld cijfer (kwantitatief) wel.

Kwantitatieve variabelen kun je weer onderverdelen in discrete variabelen en continue variabelen.

- Discrete variabelen zijn variabelen die geen tussenwaarden kunnen aannemen. Bijvoorbeeld het aantal kinderen in een gezin, een score op een toets van veertig meerkeuzevragen, leeftijd, schoenmaat, enzovoort.
- Continue variabelen zijn variabelen als lengte, gewicht, buitentemperatuur, tijd, enzovoort. Continue variabelen kunnen allerlei tussenwaarden aannemen.

Als je dataset heel groot is, kun je ervoor kiezen om de gegevens in te delen in klassen.

Deze tabel laat de lengtedata van de 154 havo 4-leerlingen in klassen verdeeld zien.

lengte	freq V	freq M
155 - < 160	6	0
160 - < 165	12	2
165 - < 170	34	3
170 - < 175	16	10
175 - < 180	10	14
180 - < 185	5	18
185 - < 190	1	11
190 - < 195	0	9
195 - < 200	1	1
200 - < 205	0	1
	85	69

Figuur 1.3

De klasse $170- < 175$ is de klasse van 170 tot 175. Dat betekent dat de lengte 170 in deze klasse zit, maar dat de lengte 175 in de volgende klasse valt, namelijk in $175- < 180$.

Opgave 4

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

Geef voor elk van deze variabelen aan of deze kwalitatief of kwantitatief is, discreet of continu en welke waarden de variabelen kunnen aannemen.

Opgave 5

Om de lengtes van de 69 jongens en 85 meisjes goed te kunnen vergelijken, maak je eerst een klassenindeling en gebruik je de relatieve frequenties.

a Maak een tabel met de relatieve frequenties bij de klassenindeling in **Uitleg 2**.

b Zet de relatieve frequenties van de lengtes van de jongens en de meisjes in afzonderlijke staafdiagrammen. Langs de horizontale as komt de lengte (cm). Langs de verticale as de relatieve frequentie (%).

Kun je op grond van deze staafdiagrammen bepalen hoeveel procent van de jongens langer is dan 182 cm? Licht je antwoord toe.

c Welke voordeel heeft het groeperen van de metingen in klassen? En welk nadeel?

d Welk nadeel heeft het vergroten van de breedte van de klassen?

Theorie en voorbeelden

Om te onthouden

Statistiek houdt zich bezig met het verzamelen, uitleggen en presenteren van betrouwbare gegevens over groepen (bijvoorbeeld mensen), de **populatie**. Meestal wordt maar een deel van de groep, de **steekproef**, onderzocht. De verzamelde gegevens heten **data**.

Als je verzamelde gegevens presenteert en gebruikt om daarin patronen te ontdekken, dan ben je bezig met **beschrijvende statistiek**. Je kunt op basis van de gegevens van een steekproef uitspraken doen over de populatie: je bent dan bezig met **verklarende statistiek**.

Het kiezen van de steekproefdeelnemers moet **aselect** plaatsvinden. Dat wil zeggen dat elk lid van de populatie even grote kans heeft om in de steekproef te komen. Daarbij wordt vaak gebruikgemaakt van **toeval**, bijvoorbeeld door loten of toevalsgetallen uit een computer of de grafische rekenmachine.

De steekproef moet **representatief** zijn. Alle soorten leden uit de populatie moeten naar verhouding even vaak in de populatie voorkomen als in de steekproef.

De **steekproefomvang** moet groot genoeg zijn. Bijvoorbeeld om te zorgen dat deze ook representatief is.

Voor het verzamelen van gegevens kan een **enquête** worden gebruikt die bestaat uit **onderzoeksvragen**.

Leeftijd, geslacht, lengte, gewicht, kleur ogen noem je **variabelen**.

Kwalitatieve variabelen zijn bijvoorbeeld geslacht en kleur ogen. Ze geven een eigenschap (kwaliteit) weer.

Kwantitatieve variabelen hebben een getalswaarde, bijvoorbeeld leeftijd, en je kunt ermee rekenen.

Er zijn verschillende typen kwantitatieve variabelen:

- **discrete** kwantitatieve variabelen, zoals schoenmaat. Alleen bepaalde waarden (bijvoorbeeld 7 en $7{,}5$) kunnen voorkomen, de andere tussenliggende waarden niet.
- **continue** kwantitatieve variabelen, zoals lengte en gewicht. Elke tussenliggende waarde kan ook voorkomen.

‘Discreet’ betekent: ‘los van elkaar’; ‘continu’ staat voor ‘aaneengesloten’.

Bekijk het **Practicum** om na te gaan hoe je **toevalsgetallen** met de grafische rekenmachine kunt oproepen.

Voorbeeld 1

Hier zie je vier manieren om een steekproef samen te stellen:

- Voor een onderzoek naar de service van de NS in de treinen ga je mensen enquêteren. Je kiest voor de uit/ingang van een treinstation en bevraagt vanaf 7:00 uur elk uur van de dag 10 willekeurige reizigers.
- Voor een onderzoek naar het rijgedrag van vrachtautochauffeurs ga je mensen enquêteren. Je kiest voor de uit/ingang van een treinstation en bevraagt vanaf 7:00 uur elk uur van de dag 10 willekeurige reizigers.
- Voor een onderzoek naar het rookgedrag van ouders van leerlingen van jullie school, ondervraag je de eerste 50 binnenkomende ouders op een ouderavond.
- Voor een onderzoek naar het rookgedrag onder ouders van leerlingen van jullie school, selecteer je door loting 50 leerlingen van jullie school en ondervraag je weer na loting de vader of de moeder van elk van de 50 leerlingen.

Welke van deze steekproeven zijn niet representatief en welke zijn niet aselect?

Antwoord

De tweede steekproef is niet representatief maar wel aselect. Je spreekt waarschijnlijk relatief weinig automobilisten. De mening van automobilisten is mogelijk sterk afwijkend van die van de treinreizigers ten aan zien van de onderzoeksvraag.

De derde steekproef is niet aselect, bijvoorbeeld ouders die tot laat werken zullen waarschijnlijk niet vroeg op de ouderavond kunnen zijn en hebben dus en kleinere kans om in de steekproef te komen. Daardoor is deze steekproef ook niet representatief.

Opgave 6

Bekijk **Voorbeeld 1**.

a Hoe kun je de tweede steekproef aselect en representatief maken?

b Over welke variabelen kan de eerste steekproef gaan? Wat voor soort variabelen betreft het?

Soms wil je dat je steekproef aan bepaalde voorwaarden voldoet, bijvoorbeeld wil je dat bepaalde leeftijdsgroepen in de werkelijke verhouding in je steekproef voorkomen. Dat heet een ‘gelaagde steekproef’.

c Hoe kun je dit bij de laatste steekproef in het voorbeeld toepassen?

Opgave 7

Naar welk soort variabele verwijst de gestelde vraag? Kies uit: kwalitatieve variabele, discrete kwantitatieve variabele of continue kwantitatieve variabele.

a Hoeveel vakken heb je?

b Hoe ver is het van school naar huis?

c Welk profiel heb je gekozen?

d Welke docent geeft je het vak Nederlands?

e Hoe lang zit je al op school?

f Hoe lang duurt het nog tot het eindexamen?

g In hoeveel vakken doe je eindexamen?

Opgave 8

In de Nationale Wetenschapsquiz kwam de volgende vraag voor. Stel je wilt weten hoeveel schoolgaande kinderen er gemiddeld per gezin zijn. Je neemt een grote steekproef onder schoolkinderen en vraagt hun hoeveel schoolgaande broertjes en zusjes zij hebben. Op basis daarvan bepaal je het gemiddelde aantal schoolgaande kinderen per gezin.

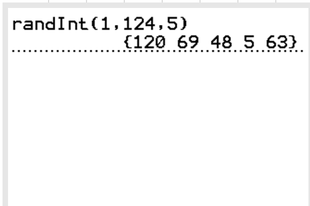
Is dit een goede aanpak? Welk van de antwoorden is correct en waarom?

A. Ja, zo krijg je een juiste schatting.
B. Nee, zo krijg je een te lage schatting.
C. Nee, zo krijg je een te hoge schatting.

Voorbeeld 2

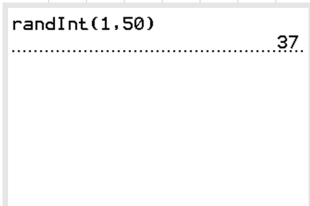
Om een steekproef samen te stellen uit een grote populatie, wordt vaak met toevalsgetallen gewerkt. Je ziet twee voorbeelden van het gebruik van toevalsgetallen. Bekijk het **Practicum** om na te gaan hoe je ze kunt oproepen met de grafische rekenmachine.

Eerste voorbeeld:
Bij een wielerwedstrijd moeten 5 renners naar de dopingcontrole. Ze hebben rugnummers vanaf 1 tot en met 124. Er worden 5 toevalsgetallen van 1 tot en met 124 gegenereerd.
De renners met rugnummers die overeenkomen met de vijf toevalsgetallen, moeten naar de dopingcontrole.



Figuur 1.4

Tweede voorbeeld:
Een op de vijftig mensen die op Schiphol door de douane wil, moet uitgebreid worden gefouilleerd en de handbagage moet worden doorzocht.
Er worden toevalsgetallen gegenereerd van 1 tot en met 50. Dit toevalsgetal is (per groep van vijftig) het nummer van de passagier die intensief wordt gecontroleerd.



Figuur 1.5

Bij statistiek gaat het om het werken met grote verzamelingen gegevens, grote datasets. Daarbij gebruik je de computer en werk je (bijvoorbeeld) met Excel.

Opgave 9

Bekijk **Voorbeeld 2**. Je ziet hoe je met behulp van toevalsgetallen een steekproef kunt samenstellen.

a Hoe kun je met behulp van toevalsgetallen uit de dagproductie van twaalfhonderd spaarlampen twintig testexemplaren kiezen?

b Hoe kun je met behulp van toevalsgetallen een steekproef van vijftienhonderd willekeurig gekozen Nederlanders samenstellen?

Soms wil je dat je steekproef aan bepaalde voorwaarden voldoet. Je wilt bijvoorbeeld dat bepaalde leeftijdsgroepen in de werkelijke verhouding in je steekproef voorkomen. Dat is een ‘gelaagde steekproef’.

c Stel je voor dat in een bepaalde stad met 60000 inwoners de percentages van de leeftijdsgroepen 0 — 20, 20 — 60 en 60 — ouder ook binnen de steekproef tot uiting komen.
Hoe kies je de personen uit die stad voor je steekproef?

Verwerken

Opgave 10

Naar welk soort variabele verwijst de gestelde vraag?

a Hoeveel vakken heb je?

A. kwalitatieve variabele
B. discrete kwantitatieve variabele
C. continue kwantitatieve variabele

b Hoe ver is het van school naar huis?

A. kwalitatieve variabele
B. discrete kwantitatieve variabele
C. continue kwantitatieve variabele

c Welk profiel heb je gekozen?

A. kwalitatieve variabele
B. discrete kwantitatieve variabele
C. continue kwantitatieve variabele

d In hoeveel vakken doe je eindexamen?

A. kwalitatieve variabele
B. discrete kwantitatieve variabele
C. continue kwantitatieve variabele

Opgave 11

Lees de onderzoeksopzet.

> ‘Belgen praten beduidend langzamer dan Nederlanders. In de Randstad haalt men 5,42 lettergrepen per seconde, in Oost-Vlaanderen slechts 4,43. Sommige mensen gingen het meteen controleren. De spreeksnelheid van 21 miljoen Nederlandssprekenden werd bepaald door maar liefst 160 leraren en leraressen een stukje te laten opzeggen. Er waren acht groepen, dus twintig sprekers per groep. Ook werd nog gerapporteerd over het verschil tussen jong en oud, man en vrouw.’
>
> **bron: tijdschrift Onze Taal, 2004, Hans van Maanen**

a Wat vind je van deze opzet?

b Wat vind je van de steekproef?

c Wat vind je van de conclusie dat Belgen beduidend langzamer praten dan Nederlanders?

d De journalist Hans van Maanen rangschikt dit onderzoek in de top 10 van wetenschappelijke blunders van 2004. Waarom denk je?

Opgave 12

In een straat staan precies honderd woningen. Het zijn twintig blokken van vijf woningen. Aan iedere kant van de weg staan tien blokken. Je hebt een even kant met de huisnummers 2 tot en met 100 en een tuin op het zuiden. Je hebt een oneven kant met de huisnummers 1 tot en met 99 en een tuin op het noorden.

a Een energiebedrijf wil het gasverbruik in deze straat onderzoeken. Het neemt een steekproef van tien huizen: de huisnummers 1, 11, 21, 31, 41, 51, 61, 71, 81 en 91. Waarom is deze steekproef niet aselect?

b Het gemiddelde gasverbruik dat de onderzoeker bij de tien huizen vindt, blijkt veel hoger te zijn dan het gemiddelde in de straat in werkelijkheid blijkt te zijn. Hoe kan dit?

c Bedenk een manier om aselect tien huizen uit de straat te selecteren voor het onderzoek, zodat het gemiddelde gasverbruik van de tien huizen representatief is voor de hele straat.

d Gebruik de rekenmachine of Excel om tien toevalsgetallen te genereren uit deze straat. En ga na of de huisnummers die je vindt representatief zijn voor de hele straat.

Opgave 13

In de jaren 1982-1988 werd onder 22000 mannelijke Amerikaanse artsen onderzoek gedaan naar de invloed van aspirine op hart- en vaatziekten op de gemiddelde Amerikaanse man. De helft gebruikte om de dag 300 milligram aspirine, wat ongeveer gelijk staat aan een ‘gewoon’ aspirientje. De andere helft slikte een placebo (‘fopmiddel’). Van de aspirineslikkers kregen 104 personen een hartinfarct, van de placeboslikkers waren dat er 189. De conclusie van het onderzoek was dat het risico op een hartinfarct met ongeveer 45% wordt verlaagd door het slikken van aspirine. Dat dit grote verschil aan toeval was te wijten, vond men uitgesloten vanwege het grote aantal mensen dat aan de studie meewerkte.

a Waarom is hier geen sprake van een representatieve steekproef? Hoe had deze steekproef moeten worden samengesteld?

b Waarom werd er gebruikgemaakt van placebo’s?

c Hoeveel procent van de 11000 aspirineslikkers heeft baat gehad bij het slikken van aspirine?

d Volgens de tekst wordt de kans op een hartinfarct met 45% verlaagd. Klopt dat?

Opgave 14

Veel onderzoek gebeurt door mensen een vragenlijst te laten invullen. Het opstellen van de juiste vragen is erg belangrijk. Op slecht gestelde vragen krijg je nutteloze antwoorden.
Stel je bent nieuwsgierig wat de leerlingen uit je klas bij het ontbijt eten.

a Je bedenkt als vraag: “Wat vind je lekkerder op de boterham, hagelslag of kaas?” Leg uit waarom deze vraag niet goed is.

b Je bedenkt ook de vraag: “Wat is gezonder: een witte boterham of een bruine boterham?” Leg uit waarom ook deze vraag niet goed is.

c Je zou ook aan iedere leerling de volgende opdracht kunnen geven: “Schrijf op wat je vanmorgen hebt gegeten als ontbijt.” Wat is een nadeel van deze vraag?

d Je zou ook aan ieder leerling de volgende opdracht kunnen geven: “Geef met een kruisje aan wat je vanmorgen als ontbijt hebt gehad.” Kies uit bruin brood, yoghurt met muesli en/of fruit. Wat is er mis met deze vraag?

e Welke vraag zou jij stellen waarop je een zinvol antwoord krijgt? Probeer uit of het echt een handige en goede vraag is.

Toepassen

Opgave 15: De Nationale Doorsnee

De ‘Nationale Doorsnee’ was in 2000 een landelijk statistiekproject voor leerlingen uit klas 1 en 2. Centrale vraag was: ‘Wie is de gemiddelde leerling van Nederland?’ Het ging bij dit project om negen kenmerken:

- lichaamslengte
- ontbijtgewoonte
- tijdsbesteding sport
- tijdsbesteding tv
- tijdsbesteding computer
- leukste vak op school
- zakgeld per week
- bijverdienste per week
- favoriete popster of popgroep

a Naar welk soort variabele verwijst elk van deze kenmerken?

b Bedenk bij elk kenmerk een goede vraag, die aansluit bij de door jou genoemde soort variabele.

c Welk tiende kenmerk en welke tiende vraag zou je kunnen toevoegen om de gemiddelde leerling van Nederland nog verder te typeren?

Opgave 16: Smartphonegebruik

Maak een opzet voor een onderzoek onder een deel van de ouders en leerlingen van de school naar hun gebruik van een smartphone en neem je onderzoeksplannen onder de loep en verbeter deze.

a Ontwerp minimaal twee heel concrete onderzoeksvragen: wat wil je precies weten over het smartphone-gebruik van de leerlingen en hun ouders? Anders gezegd: welke vragen wil je beantwoord zien nadat je dit onderzoek hebt uitgevoerd?

b Ontwerp een lijst met goede vragen, zodanig dat je met de antwoorden erop je eigen onderzoeksvragen kunt beantwoorden.

c Stel een lijst met variabelen samen waarin je de antwoorden op je vragenlijst gaat vastleggen en geef bij elke variabele aan of het om een kwalitatieve, discreet kwantitatieve of continu kwantitatieve variabele gaat.

d Ontwerp een aselecte en representatieve steekproef van leerlingen en ouders op jouw school.

e Ontwerp de manier waarop je de steekproefpersonen de vragen gaat stellen zodanig dat je de meeste kans hebt op zo veel mogelijk serieuze antwoorden.

Testen

Opgave 17

In 2006 hield het mannenblad ‘Men’s Health’ een wereldwijd seksonderzoek. Ruim 40000 mannen (lezers) van het blad vulden de vragenlijst in. In Nederland waren er ongeveer 1500 mannen die reageerden. De kop boven veel artikelen was: ‘Nederlandse man vrijt langdurig.’

a Wat vind je van de opzet van het onderzoek?

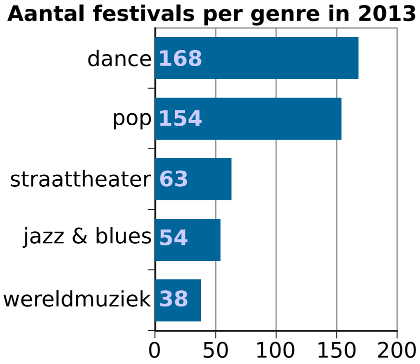
b Is de steekproef aselect en representatief?

c Wat vind je van de kop boven de artikelen?

d Wat is het lastige van vragen bij seksonderzoeken en de conclusies die je eruit kunt trekken?

Opgave 18

In deze afbeelding kun je voor ieder genoemd muziekgenre het aantal muziekfestivals dat in Nederland in 2013 plaatsvond, aflezen.



Figuur 1.6

a Van welk type variabele worden hier gegevens afgebeeld: van een discrete kwantitatieve variabele of van een kwalitatieve variabele?

Deze gegevens zijn verzameld vanwege een statistisch onderzoek.

b Bedenk een mogelijk onderzoek waarbij deze gegevens als onderdeel van een aselecte, representatieve steekproef worden gebruikt. Omschrijf duidelijk de onderzoeksvraag, de populatie en de steekproef.

c Bedenk een mogelijk onderzoek waarbij dit de gegevens van de gehele populatie zijn. Omschrijf duidelijk de onderzoeksvraag en de populatie.

Stel dat deze gegevens gebruikt worden voor het beantwoorden van de onderzoeksvraag ‘Welk genre muziekfestival kwam het meest voor in Nederland in 2013?’.

d Is dit dan een vorm van beschrijvende of verklarende statistiek? Verklaar je antwoord.

Opgave 19

De docent van een bovenbouwgroep wil aselect een leerling kiezen als assistent bij de uitleg. In de groep zitten 32 leerlingen.

a Bedenk een manier om dit met een getalsimulatie op de grafische rekenmachine of in (een digitaal programma zoals) Excel uit te voeren.

b Bedenk een manier om dit met een (getal-)simulatie zonder rekenmachine of digitaal hulpmiddel uit te voeren.

c Bedenk een manier om zonder simulatie, hopelijk aselect, een leerling uit deze groep te kiezen.

Practicum

Met de volgende practica kun je leren hoe je **toevalsgetallen** met de grafische rekenmachine genereert.

- **Simulaties en de TI84**
- **Simulaties en de TInspire**
- **Simulaties en de Casio fx-CG50**
- **Simulaties en de HPprime**
- **Simulaties en de NumWorks**

Met Excel (een spreadsheetprogramma, een rekenblad) werken is bij statistiek eigenlijk onontbeerlijk. Je kunt er grote hoeveelheden gegevens in kwijt. Bekijk deze practica voor **Excel 2013/2016/2019**:

- **Tafels** om de basisbeginselen van het werken met Excel te leren.
- **Diagrammen** om te leren hoe je in Excel lijn-, staaf-, cirkeldiagrammen kunt maken.
- **Data presenteren en vergelijken** om te bekijken hoe je grote databestanden kunt samenvatten en deelgroepen daarin kunt vergelijken.
- Van **Steekproeven en uitspraken**, alleen het eerste deel: “Een steekproef trekken met toevalsgetallen”.

2.2 Ordenen

Inleiding

Een onderzoek levert veel waarnemingen (antwoorden, kenmerken of meetresultaten) op. Je zult eerst overzicht moeten krijgen over al die gegevens door ze te ordenen. Dat doe je door bijvoorbeeld ‘turven’ hoe vaak een antwoord voor komt.

Figuur 2.1

Je leert in dit onderwerp

- de ruwe gegevens uit een onderzoek te ordenen in een overzichtelijke tabel;
- bij veel verschillende gegevens een klassenindeling te maken en de begrippen klassenbreedte en klassengrens;
- de begrippen turftabel, frequentietabel en somfrequentietabel en het onderscheid tussen frequentie en relatieve frequentie;
- het begrip proportie.

Voorkennis

- de begrippen steekproef en populatie;
- rekenen met procenten.

Verkennen

Opgave V1

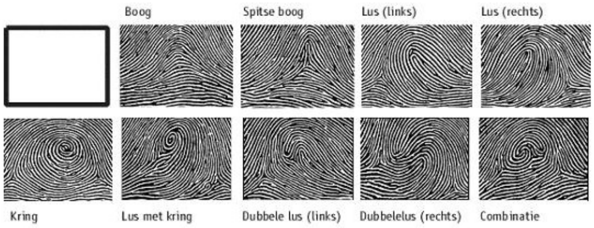
Sir Francis Galton heeft rond 1890 vingerafdrukken bestudeerd. Hij ontdekte dat je bij vingerafdrukken grofweg drie patronen kunt onderscheiden: de boog, de kring en de lus.
Deze indeling is vrij globaal. Voor sommige patronen moeten ook nog het aantal lijnen tussen de kern en delta geteld worden. Hierdoor is ieder mens te identificeren op grond van de vingerafdruk.

Hoe zou je gegevens over vingerafdrukken van bijvoorbeeld alle leerlingen kunnen verzamelen, ordenen en snel kunnen terugvinden in een bestand?

Uitleg

Een statistisch onderzoek levert bijvoorbeeld antwoorden, waarnemingen, kenmerken of meetresultaten. Je moet eerst overzicht krijgen over al die gegevens met sorteren en samenvatten. Dat is beschrijvende statistiek.

Sir Francis Galton heeft rond 1890 vingerafdrukken bestudeerd. Hij ontdekte dat je bij vingerafdrukken grofweg drie patronen kunt onderscheiden: de boog, de kring en de lus.
Deze indeling is vrij globaal. Voor sommige patronen moeten ook nog het aantal lijnen tussen de kern en delta geteld worden. Hierdoor is ieder mens te identificeren op grond van de vingerafdruk.



Figuur 2.2

Je ziet hoe door 'turven' een frequentietabel van kenmerk 'patroon van linker duimafdruk' ontstaat voor een groep van 25 personen. Er is gebruikgemaakt van de verdeling in drie hoofdcategorieën van vingerafdrukken. De relatieve frequentie, ook wel 'proportie' genoemd, ontstaat door de absolute frequentie te delen door het totaal aantal waarnemingen.

linker duimafdruk	aantal	abs.freq.	rel.freq.	percentage
lus	||||| |||	8	$\frac{8}{25} = 0{,}32$	32%
boog	||||| |||||		11	$\frac{11}{25} = 0{,}44$	44%
kring	|||||		6	$\frac{6}{25} = 0{,}24$	24%

Tabel 2.1

Bij een tabel met (relatieve) frequenties kun je een tabel met (relatieve) somfrequenties maken door bij elke frequentie de voorafgaande frequenties op te tellen. Somfrequenties noem je cumulatieve frequenties (cumuleren betekent opstapelen).

Bij continue variabelen moet je gebruikmaken van een klassenindeling. Zorg dat je ongeveer tien klassen krijgt om mee te werken.

Stel bijvoorbeeld dat je de lengtes van een groep meisjes onderzoekt. Is het kleinste meisje $1{,}51$ m en het langste $1{,}98$ m, dan kun je klassen maken met een klassenbreedte van 5 cm. De klassen sluiten altijd op elkaar aan. De eerste klasse is $1{,}50- < 1{,}55$. De tweede klasse is $1{,}55- < 160$, enzovoort.

De notatie $- <$ betekent 'vanaf ... tot ...'. De waarden $1{,}50$ en $1{,}55$ van de eerste klasse worden de klassengrenzen genoemd. Een meisje dat afgerond $1{,}55$ m lang is, zit in de tweede klasse, want die klasse begint bij $1{,}55$.

Opgave 1

Bestudeer de **Uitleg**.
Bekijk de tabel met daarin de lengtes van twintig meisjes. Je gaat hierbij een frequentieverdeling maken met klassenindeling.

1,63	1,72	1,86	1,66
1,73	1,92	1,66	1,74
1,95	1,68	1,76	1,53
1,70	1,78	1,55	1,71
1,80	1,81	1,72	1,83

Tabel 2.2

a Waarom gebruik je hier een klassenindeling?

b Je gaat tien klassen maken. Wat wordt dan de eerste klasse?

c Maak de frequentietabel met een klassenindeling en reken in de frequentietabel alle absolute frequenties om via de fractie naar relatieve frequenties en absolute frequenties.

d Welke proportie hoort bij $1{,}70- < 1{,}75$?

e Maak zo'n relatieve frequentietabel voor de lengtes van alle leerlingen in je klas. Welke klassengrenzen kies je?

Opgave 2

Bekijk de klassenindeling met daarin de lengtes van twintig meisjes.

	aantal	proportie	rel. freq.
1,50− < 1,70	6	0,3	30%
1,70− < 1,75	6	0,3	30%
1,75− < 1,85	6	0,5	30%
1,90− < 1,95	1	0,05	5%
1,95− < 2,00	1	0,05	5%
Totaal	20	1	100%

Tabel 2.3

a Wat valt je op aan de klassenindelingen?

b Welke eerste indruk wekt deze klassenindeling?

c Leg uit waarom dergelijke klassenindelingen niet goed bruikbaar zijn.

Opgave 3

Bedenk zelf iets wat je wilt onderzoeken. Maak de opzet van je onderzoek zo dat je het ook echt kunt uitvoeren.

Volg de volgende stappen:

1. Ontwerp minimaal twee heel concrete onderzoeksvragen: welke vragen wil je beantwoord zien nadat je dit onderzoek hebt uitgevoerd?
2. Ontwerp een lijst met goede vragen, zodanig dat je met de antwoorden erop je eigen onderzoeksvragen kunt beantwoorden.
3. Stel een lijst met variabelen samen waarin je de antwoorden op je vragenlijst gaat vastleggen en geef bij iedere variabele aan of het om een kwalitatieve, discreet kwantitatieve of continu kwantitatieve variabele gaat.
4. Ontwerp een aselecte en representatieve steekproef.
5. Ontwerp de manier waarop je de steekproefpersonen de vragen gaat stellen zodanig dat je de meeste kans hebt op zo veel mogelijk serieuze antwoorden.
6. Voer je onderzoek uit.
7. Geef je resultaten weer in tabellen. Bedenk bij elke tabel of een klassenindeling wel of niet handig is.

Let op! Je hoeft nog geen diagrammen te maken en/of uitspraken te doen. Bewaar je resultaten wel goed, je gaat hier in de volgende paragrafen mee verder.

Tabel 2.4

Theorie en voorbeelden

Om te onthouden

De **absolute frequentie** is het aantal keren dat een waarneming voorkomt. Het bijhouden van het aantal keren dat een waarneming voorkomt, gebeurt vaak in een **turftabel**. Een **frequentietabel** is het overzicht van het aantal keren dat alle waarnemingen voorkomen. Dit is een **frequentieverdeling**.

De **proportie** is de frequentie van een waarneming gedeeld door het totale aantal waarnemingen. Dit is een **relatieve frequentie**.

Bij een frequentietabel kun je een **somfrequentietabel** maken door bij elke frequentie de voorafgaande frequenties op te tellen. Zo maak je ook een relatieve somfrequentietabel. Somfrequenties noem je **cumulatieve frequenties** (cumuleren betekent opstapelen).

Heb je veel verschillende waarnemingsgetallen, dan werk je met een **klassenindeling**. Klassen kun je op twee manieren noteren.

- vanaf $1{,}50$ tot $1{,}60$ wordt: $1{,}50- < 1{,}60$ ($1{,}60$ zelf zit in de volgende klasse)
- vanaf 10 tot en met 14 wordt: $10 - 14$ (in deze klasse zit zowel 10 als 14)

Voor continue kwantitatieve variabelen gebruik je de eerste notatie. Voor discrete kwantitatieve variabelen gebruik je de tweede notatie. Voor leeftijden (alleen hele getallen) is dit lastiger, want $10 - 14$ betekent dan $10- < 15$.

De **klassengrenzen** zijn de onderste en bovenste waarnemingen van een klasse. Het verschil tussen twee opvolgende klassengrenzen is de **klassenbreedte**. Neem met betrekking tot het opstellen van klassen de volgende regels in acht:

- Neem voor het aantal klassen minimaal vijf en maximaal twintig. Maak het aantal (ongeveer) gelijk aan de wortel uit het aantal waarnemingsgetallen.
- Voor de klassenbreedte neem je het verschil tussen de grootste en de kleinste waarneming en dat deel je door het aantal klassen. Vaak rond je af op een geheel getal.

Het **klassenmidden** gebruik je voor het schatten van het gemiddelde van een frequentieverdeling.
Gaat het om de variabele leeftijd (discrete variabele), dan betekent $5 - 9$ hetzelfde als $5- < 10$. Het is de klasse met de leeftijden 5, 6, 7, 8 en 9. Het klassenmidden is $7{,}5$.
Bij bijvoorbeeld de variabele lengte (continue variabele) wordt met de notatie $140- < 145$ de klasse bedoeld van 140 tot 145. Omdat in deze klasse alle waarden tot 145 zitten, is het klassenmidden $142{,}5$.
Bij grote datasets ontkom je er niet aan om met klassenindelingen te werken.

Voorbeeld 1

Bekijk de tabel met de lengtes in cm van negentig meisjes.

175	168	177	167	176	172	166	160	166
173	172	170	186	168	165	159	164	183
155	179	184	155	188	163	156	172	163
161	162	174	159	162	169	171	179	170
170	165	157	168	167	166	172	174	158
183	173	168	150	182	154	160	159	168
189	153	162	166	157	179	164	169	175
165	193	154	180	171	168	180	181	173
171	176	165	176	172	169	161	167	165
159	169	176	185	176	164	169	166	165

Tabel 2.5

Maak een klassenindeling met een klassenbreedte van 5 cm en als ondergrens $1{,}50$ m. Bereken ook de relatieve frequenties.

Antwoord

Maak de klassenindeling en turf de aantallen; maak een (absolute) frequentietabel. Excel kan automatisch turven in series getallen. In het **Practicum** kun je zien hoe dat in zijn werk gaat. Let erop dat de eerste klasse $1{,}50- < 1{,}55$ is, de tweede klasse $1{,}55- < 1{,}60$, enzovoort.

klasse	abs.freq.	rel.freq. (%)
$150- < 155$	4	4,4
$155- < 160$	10	11,1
$160- < 165$	12	13,3
$165- < 170$	25	27,8
$170- < 175$	16	17,8
$175- < 180$	11	12,2
$180- < 185$	7	7,8
$185- < 190$	4	4,4
$190- < 195$	1	1,1

Tabel 2.6

Opgave 4

Gebruik de tabel met de lengtes in cm van negentig meisjes in **Voorbeeld 1**.

a Verander de klassenbreedte van 5 in 8. Maak een tabel met de klassen, absolute frequenties, relatieve frequenties en de cumulatieve frequenties.

b Verander de klassenbreedte van 5 naar 2. Hoe groot is nu de hoogste frequentie?

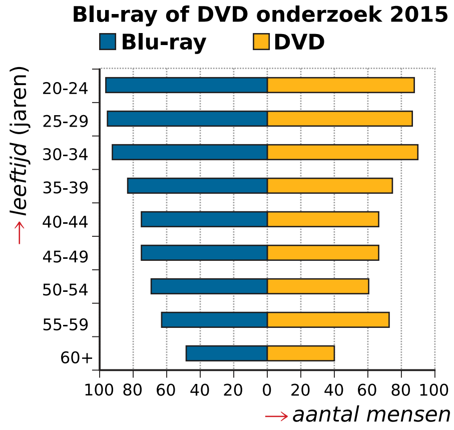
c Probeer nog een aantal klassenbreedten. Welke klassenbreedte vind je het meest geschikt?

Voorbeeld 2

In dit leeftijdsdiagram van een IT-bedrijf in 2015 zie je hoeveel mensen per leeftijdscategorie de voorkeur aan Blu-ray geven en hoeveel aan dvd.

De klasse $20-24$ bijvoorbeeld bevat werknemers die een leeftijd hebben vanaf 20 tot 25 jaar.

Kun je met de gegevens in dit diagram een nieuw leeftijdsdiagram maken met klassen van $20-27$, $28-35$, $36-43$, enzovoort? En met klassen van $20-29$, $30-39$, enzovoort? Licht je antwoord toe.



Figuur 2.3

Antwoord

Omdat in de klasse $20-27$ de klasse $20-24$ geheel en de klasse $25-29$ voor een deel zit, kun je uit dit diagram niet opmaken hoeveel werknemers uit de klasse $25-29$ bij de klasse $20-27$ moeten komen. Je kent de onderverdeling van de klassen namelijk niet. Je weet alleen het totale aantal werknemers in de gegeven klassen. Voor de klassen $20-29$, $30-39$, is dat anders omdat je nu het aantal werknemers uit twee gegeven klassen bij elkaar op kunt tellen. Met de klassenindeling $20-29$, $30-39$ ... $50-59$ zou je dus wel een nieuw diagram kunnen tekenen.

Opgave 5

Bekijk **Voorbeeld 2**. Je ziet een leeftijdsdiagram van een IT-bedrijf in 2015, hierin wordt de voorkeur voor Blu-ray of dvd aangegeven.

a Welke klassengrenzen heeft de klasse $35-39$?

b Als je het aantal klassen van 9 in 3 verandert, hoe groot is dan de hoogste frequentie bij Blu-ray?

c Waarom is het verhogen van het aantal klassen nu niet mogelijk zonder extra informatie?

Opgave 6

Welke van de volgende beweringen zijn juist?

A. In een relatieve frequentietabel of relatieve somfrequentietabel staan altijd percentages.

B. De totale relatieve somfrequentie is in theorie altijd 100%.

C. De totale relatieve somfrequentie is in de praktijk altijd 100%.

D. De relatieve frequentie is overal 100%.

E. Als er waarnemingen in de laatste klasse vallen, zijn de relatieve somfrequenties lager dan 100%, behalve bij de laatste klasse.

Verwerken

Opgave 7

Je ziet de frequentieverdeling van de weekklonen van 65 werknemers van een bedrijf.

loon (€)	aantal
500– < 600	8
600– < 700	10
700– < 800	16
800– < 900	14
900– < 1000	10
1000– < 1100	5
1100– < 1200	2
totaal	65

Tabel 2.7

a De frequenties in deze tabel zijn

A. absoluut
B. relatief

b Is de indeling in klassen goed?

A. ja
B. nee

c De klassenbreedte in deze frequentietabel is

A. 99,00
B. 99,99
C. 100,00

d Welke uitspraken zijn juist?

A. De relatieve frequentie van lonen tussen de € 700,00 en € 800,00 is 16%.
B. De relatieve frequentie van lonen tussen de € 700,00 en € 800,00 is 24,6%.
C. De proportie lonen van minstens € 1000,00 is ongeveer 0,11.
D. De cumulatieve relatieve frequentie van lonen minder dan € 700,00 is 18%.

Opgave 8

Voor een toets kun je maximaal 100 punten scoren. Je ziet hoe een groep van veertig 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

a Deel deze scores in klassen in. Neem als laagste klasse 25– < 35. Maak een frequentietabel.

b Maak bij deze tabel een kolom van relatieve frequenties.

Opgave 9

De downloadsnelheid is afhankelijk van de connectiesnelheid, van de internetprovider, van welke computer/server gedownload wordt en van de drukte op het net. Zo kan de downloadsnelheid op het ene moment 72 kB/s (kilobyte per seconde) zijn bij een connectiesnelheid van 590 kbit/s en op een ander moment 155 kB/s bij een connectiesnelheid van 1270 kbit/s, terwijl de provider 3072 kbit/s (3,1 Mbit/s) ofwel 384 kB/s opgeeft. (Een byte bestaat uit acht bits. Om van kbit/s naar kB/s te gaan, moet er door acht gedeeld worden.)

Justin heeft tijdens het downloaden van een computerprogramma elke minuut de downloadsnelheid opgeschreven. De resultaten zie je in de tabel, deze staan in kB/s.

380	376	366	340	345
330	354	321	299	290
287	301	340	330	344
369	384	388	376	377
356	361	344	351	365
375	384	390	388	406

Tabel 2.8

a Hoeveel klassen kun je het beste maken voor deze gegevens?

b Hoe groot wordt dan de klassenbreedte?

c Schrijf de laagste klasse op.

d Deel deze scores in klassen in. Maak een frequentietabel.

e Maak bij deze tabel een kolom van relatieve frequenties. Let er bij het afronden op, dat het totaal 100% blijft.

f Maak bij deze tabel een kolom van relatieve cumulatieve frequenties. Let er bij het afronden op dat het totaal 100% blijft.

Opgave 10

Genereer in Excel of met de grafische rekenmachine honderd toevalsgetallen van 1 tot en met 20.

a Maak een turftabel.

b Maak een frequentietabel.

c Maak een tabel met relatieve frequenties en somfrequenties.

d Welke relatieve frequenties verwacht je bij de twintig getallen als je 10^6 toevalsgetallen van 1 tot en met 20 zou genereren?

Opgave 11

Bekijk de frequentietabellen met weeklonen van twee bedrijven. Alle werknemers zijn opgenomen in de tabellen.

weekloon (euro)	*aantal werknemers*
500– < 600	8
600– < 700	10
700– < 800	16
800– < 900	14
900– < 1000	10
1000– < 1100	5
1100– < 1200	2
totaal	65

Bedrijf 1

weekloon (euro)	*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
totaal	25

Bedrijf 2

Tabel 2.9

a Noem twee redenen waarom je de weeklonen van deze twee bedrijven niet zinvol met elkaar kunt vergelijken als je alleen naar deze frequentietabellen kijkt.

b Maak frequentietabellen waarmee je de weeklonen van deze twee bedrijven wel goed kunt vergelijken.

c Een van de onderzoeksvragen is: in welk bedrijf zijn er relatief meer mensen die minder dan € 600,00 per week verdienen?

Uit welk soort frequentietabel zou je dit direct kunnen aflezen? Geef een antwoord op deze onderzoeksvraag.

d Het is niet mogelijk om de percentages werknemers die minder dan € 650,00 per week verdienen met elkaar te vergelijken. Leg uit waarom dat niet kan en bedenk een manier om daar wel een schatting van te kunnen maken.

Toepassen

Opgave 12: Het CBS

Het **Centraal Bureau voor de Statistiek (CBS)** houdt veel geordende statistieken bij die van belang zijn voor Nederland en de Nederlandse overheid.

Het CBS doet regelmatig onderzoek naar de opbouw van de Nederlandse bevolking met betrekking tot geslacht, leeftijd, burgelijke staat, e.d. Je ziet hier een dataset vanuit Statline, gemaakt in januari 2020. De variabele 'Groene druk' stelt de verhouding voor tussen het aantal 0 tot 20 jarigen en het aantal 20 tot 65 jarigen. De variabele 'Grijze druk' stelt de verhouding voor tussen het aantal 65-plussers en het aantal 20 tot 65 jarigen.

Figuur 2.4

Onderwerp		1900	1950	2000	2010	2017
Bevolking op 1 januari						
Naar geslacht						
Mannen en vrouwen	**x 1 000**	5 104	10 027	15 864	16 575	17 082
Mannen	**x 1 000**	2 521	4 998	7 846	8 203	8 475
Vrouwen	**x 1 000**	2 583	5 029	8 018	8 372	8 606
Naar leeftijd						
Totaal bevolking	**x 1 000**	5 104	10 027	15 864	16 575	17 082
0 tot 20 jaar	**x 1 000**	2 264	3 742	3 873	3 928	3 817
20 tot 45 jaar	**x 1 000**	1 732	3 597	5 976	5 490	5 281
45 tot 65 jaar	**x 1 000**	802	1 916	3 863	4 619	4 824
65 tot 80 jaar	**x 1 000**	272	671	1 652	1 890	2 395
80 jaar of ouder	**x 1 000**	35	100	500	648	764
Demografische druk						
Groene druk	**%**	89,3	67,9	39,4	38,9	37,8
Grijze druk	**%**	12,1	14,0	21,9	25,1	31,3

Figuur 2.5

a Welke rijen van deze tabel bevatten echte data en welke rijen worden daaruit afgeleid?

b Over welke soorten variabelen gaan deze tabellen? Waarom kiest het CBS daar voor?

c Waarom is het CBS geïnteresseerd in de variabele 'Grijze druk'?

d Bereken zelf de waarden van de 'Grijze druk' en leg uit wat het oplopen van dat getal voor de overheid betekent.

e Welke betekenis heeft de 'Groene druk'?

f Hoe kun je zien dat in NL de proportie mannen kleiner is dan de proportie vrouwen?
Hoeveel bedraagt de proportie mannen in 2017?

Opgave 13: Aantal geslaagden havo en vwo

Deze tabel laat het aantal geslaagden zien op havo en vwo gedurende een drietal schooljaren.

Voortgezet onderwijs; geslaagden naar onderwijssoort									
Onderwerpen	Aantal examenkandidaten			Aantal geslaagden			Percentage geslaagden		
Perioden	2005/'06	2006/'07	2007/'08*	2005/'06	2006/'07	2007/'08*	2005/'06	2006/'07	2007/'08*
Onderwijssoort	aantal	aantal	aantal	aantal	aantal	aantal	%	%	%
Vwo	31633	32961	35150	29287	30331	32398	93	92	92
Vwo natuurprofiel totaal	14967	15940	17233	13750	14570	15723	92	91	91
Vwo natuur en techniek	3765	3510	3970	3476	3213	3662	92	92	92
Vwo natuur en gezondheid	9641	10268	11259	8771	9275	10097	91	90	90
Vwo natuur ongedeeld	1561	2162	2004	1503	2082	1964	96	96	98
Vwo maatschappijprofiel totaal	16657	17007	17908	15528	15747	16666	93	93	93
Vwo economie en maatschappij	10087	10142	10801	9413	9392	10112	93	93	94
Vwo cultuur en maatschappij	6359	6623	6858	5911	6126	6309	93	92	92
Vwo maatschappij ongedeeld	211	242	249	204	229	245	97	95	98
Havo	42983	44560	46313	38105	39755	41371	89	89	89
Havo natuurprofiel totaal	12008	12641	13634	10743	11181	12306	89	88	90
Havo natuur en techniek	3699	3618	3871	3319	3203	3492	90	89	90
Havo natuur en gezondheid	7360	7853	8704	6546	6906	7815	89	88	90
Havo natuur ongedeeld	949	1170	1059	878	1072	999	93	92	94
Havo maatschappijprofiel totaal	30974	31911	32679	27361	28571	29065	88	90	89
Havo economie en maatschappij	15705	16203	16959	13607	14321	14993	87	88	88
Havo cultuur en maatschappij	14751	15182	15264	13294	13778	13649	90	91	89
Havo maatschappij ongedeeld	518	526	456	460	472	423	89	90	93

Figuur 2.6

a Welke variabelen worden er onderzocht? Geef per variabele aan of die kwalitatief of kwantitatief is.

b Op het havo is in 2007/2008 het aantal leerlingen dat geslaagd is 89%. Toon dit aan door een berekening.

c Hoeveel procent van de examenkandidaten havo heeft in 2007/2008 een N-profiel gekozen? Rond af op één decimaal.

d Kun je het percentage geslaagden havo in 2007/2008 berekenen vanuit de percentages geslaagden van de M-profielen en de N-profielen afzonderlijk?

e Het aantal examenkandidaten met een N-profiel op havo neemt absoluut gezien toe. Is dat relatief bekeken ook zo?

f Kun je verklaren waarom de examenkandidaten in alle ongedeelde N-profielen of M-profielen een hoger geslaagdenpercentage hebben dan het totaal per profiel?

(bron: CBS)

Testen

Opgave 14

Voor een biologiepracticum moet het aantal slakken op een stuk grond worden geteld. Het stuk grond wordt in stukken van 1 m^2 verdeeld. Iedere leerling telt het aantal slakken op vier van die stukken. Je ziet de resultaten.

aantal slakken per m^2	2	3	4	5	6	7	8	9
frequentie	16	14	7	4	2	3	1	1

Tabel 2.10

a Benoem de populatie, de variabele, en het soort variabele.

b Hoeveel m^2 is de oppervlakte van het stuk grond?

c Hoeveel leerlingen hebben er geteld?

d Hoeveel slakken zijn er totaal geteld?

e Hoeveel bedraagt de proportie m^2 met 7 slakken op het stuk grond?

f Hoeveel slakken zijn er gemiddeld per m^2 gevonden?

Opgave 15

Leerlingen in een brugklas hebben hun schoenmaat gegeven.
40 - 42 - 37 - 38 - 40
35 - 41 - 36 - 38 - 37
38 - 40 - 40 - 40 - 39
40 - 39 - 38 - 41 - 40
41 - 39 - 39 - 39 - 34
41 - 37 - 38 - 45 - 42

a Maak een frequentietabel en een tabel met relatieve frequenties.

b Maak ook een cumulatieve relatieve frequentietabel van de schoenmaten.

c Hoeveel procent van de leerlingen in deze klas heeft een schoenmaat boven de 40?

d Maak een frequentietabel met relatieve frequenties en cumulatieve relatieve frequenties waarin de schoenmaten zijn opgedeeld in klassen van vier schoenmaten breed.

e Maak op basis van deze nieuwe frequentietabel een schatting van het percentage leerlingen in deze klas met een schoenmaat boven de 40. Waarom is dat alleen een schatting?

Practicum

Met de volgende practica kun je leren hoe je **toevalsgetallen** met de grafische rekenmachine genereert.

- **Simulaties en de TI84**
- **Simulaties en de TInspire**
- **Simulaties en de Casio fx-CG50**
- **Simulaties en de HPprime**
- **Simulaties en de NumWorks**

Met Excel (een spreadsheetprogramma, een rekenblad) werken is bij statistiek eigenlijk onontbeerlijk. Je kunt er grote hoeveelheden gegevens in kwijt. Bekijk deze practica voor **Excel 2013/2016/2019**:

- **Tafels** om de basisbeginselen van het werken met Excel te leren.
- **Diagrammen** om te leren hoe je in Excel lijn-, staaf-, cirkeldiagrammen kunt maken.
- **Data presenteren en vergelijken** om te bekijken hoe je grote databestanden kunt samenvatten en deelgroepen daarin kunt vergelijken.
- Van **Steekproeven en uitspraken**, alleen het eerste deel: “Een steekproef trekken met toevalsgetallen”.

2.3 Diagrammen

Inleiding

Nadat de gegevens geordend zijn in een tabel, kun je deze ook grafisch verwerken. Een figuur is vaak gemakkelijker te lezen en is toegankelijker voor de lezer. Maar welk type diagram kun je nu het beste gebruiken? Het begrip **diagram** is een zeer algemeen begrip; er is dus heel veel keuze.

Figuur 3.1

Je leert in dit onderwerp

- verschillende soorten diagrammen te tekenen en af te lezen: staafdiagram, histogram, cirkeldiagram, steelbladdiagram, lijndiagram, frequentiepolygoon, cumulatief frequentiepolygoon, Lorenzcurve;
- een keuze te maken uit de genoemde diagrammen om je onderzoeksgegevens te verwerken.

Voorkennis

- statistische begrippen, zoals steekproef, absolute, relatieve en somfrequentie en proportie;
- gegevens in een frequentietabel (zowel met absolute als met relatieve frequenties) verwerken.

Verkennen

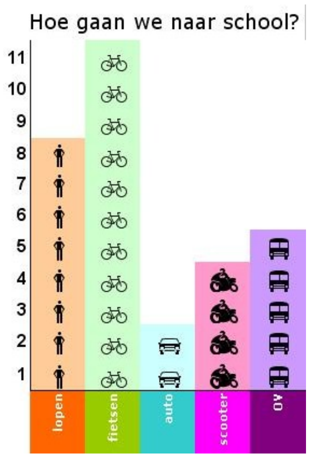
Opgave V1

Zoek op internet naar afbeeldingen van diagrammen. Mogelijke zoekwoorden zijn: diagram, chart of graph (voor Engelstalige sites). Geef bij elk diagram aan welk verhaal het vertelt; je vindt daar vaak informatie over in de kop van het betreffende artikel.

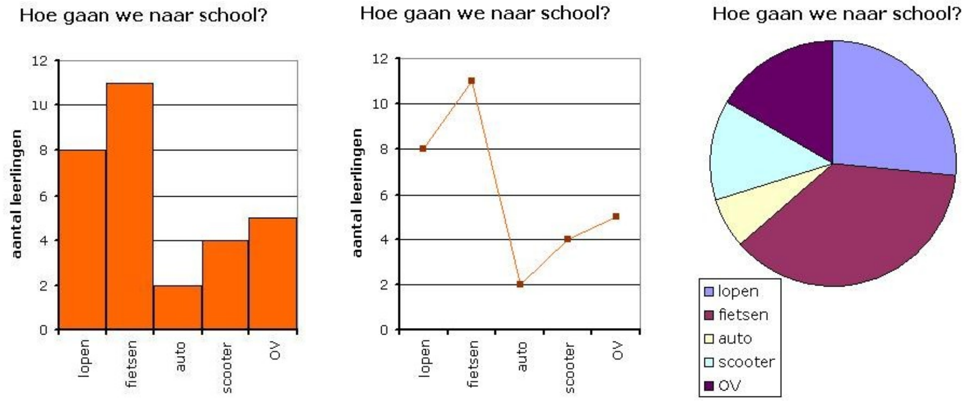
Uitleg

Een diagram is een grafische voorstelling van de (relatieve) frequenties van een statistische variabele.

Dit beelddiagram laat de frequenties van de kwalitatieve variabele *vervoersmiddel* zien. Deze variabele zie je verder nog in de vorm van een staafdiagram, een lijndiagram en een cirkeldiagram. Bij het maken van een cirkeldiagram reken je de frequenties om naar een sectorhoek. Dat doe je door de relatieve frequentie te vermenigvuldigen met 360°.



Figuur 3.2 beelddiagram



staafdiagram lijndiagram cirkeldiagram

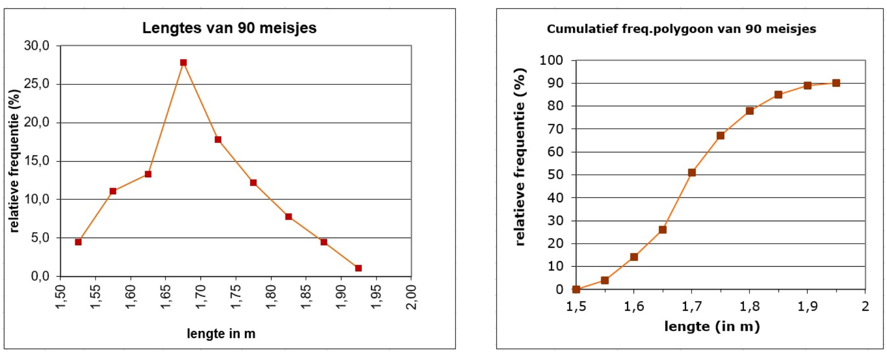
Figuur 3.3

Een histogram is een bijzonder staafdiagram. Je gebruikt het histogram alleen voor een continue kwantitatieve variabele. Bijvoorbeeld bij de lengtes van meisjes uit 4 havo. De horizontale as is dan een getallenlijn.

Een lijndiagram of frequentiepolygoon (polygoon = veelhoekig) ontstaat door in een histogram de middens van de bovenkanten van de staven te verbinden met lijnstukken en daarna de staven te verwijderen.

Een cumulatief frequentiepolygoon ontstaat uit een histogram van somfrequenties. Daarvoor verbind je de rechterbovenkanten van de staven.

Bij een diagram van een continue variabele en bij een diagram met een klassenindeling zet je de klassengrenzen links en rechts van de punt of de staaf. Anders (dus bij een discrete variabele zonder klassenindeling) staan de waarnemingsgetallen midden onder de punten of staven.



Figuur 3.4

Opgave 1

Bekijk in de **Uitleg** de vier verschillende soorten diagrammen met de manier waarop jullie naar school gaan.

a Maak met het beelddiagram een frequentietabel.

b Maak bij de frequentietabel een kolom met relatieve frequenties.

c Maak een staafdiagram en een frequentiepolygoon met relatieve frequenties.

d Bereken de sectorhoeken van het cirkeldiagram.

e Welk voordeel hebben relatieve frequenties boven absolute frequenties?

f Doe zelf een onderzoekje naar de manier waarop je klasgenoten naar school gaan.

Opgave 2

Bekijk het frequentiepolygoon van de lengtes van negentig meisjes in de **Uitleg**.

a Welke klassenindeling is gebruikt? En welke klassenbreedte?

b Hoe kun je dit frequentiepolygoon omzetten naar een histogram?

c Hoe maak je een cumulatief frequentiepolygoon?

Je vindt de gegevens uit de tabel in het bestand **Lengtes van 90 meisjes**.

d Maak een relatief frequentiepolygoon met dezelfde klassenindeling als in de uitleg.

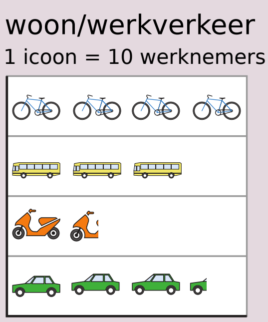
e Maak ook het cumulatief frequentiepolygoon.

Lengtes van 90 meisjes								
175	168	177	167	176	172	166	160	166
173	172	170	186	168	165	159	164	183
155	179	184	155	188	163	156	172	163
161	162	174	159	162	169	171	179	170
170	165	157	168	167	166	172	174	158
183	173	168	150	182	154	160	159	168
189	153	162	166	157	179	164	169	175
165	193	154	180	171	168	180	181	173
171	176	165	176	172	169	161	167	165
159	169	176	185	176	164	169	166	165

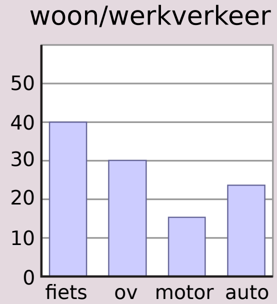
Figuur 3.5

Theorie en voorbeelden

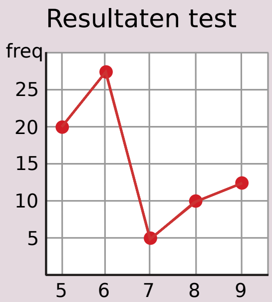
Om te onthouden



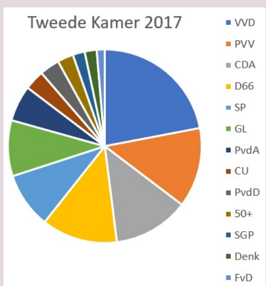
beelddiagram



staafdiagram



lijndiagram

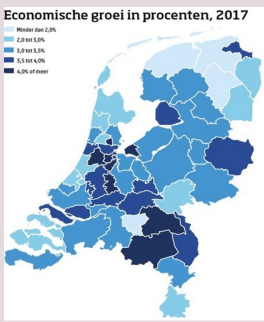


cirkeldiagram

snelheidsmeting binnen de bebouwde kom

	auto's km/h					
4	1	3	5			
5	0	3	5	6	6	9
6	1	3				
7	5					

steelbladdiagram



kaartdiagram

Figuur 3.6

Bij de analyse en presentatie van onderzoeksgegevens kun je vaak gebruikmaken van visuele weergaves. Een **diagram** is een grafische voorstelling van de (relatieve) frequenties van een statistische variabele.

Een **staafdiagram** en een **beelddiagram** geven de (relatieve) frequentie weer als staafhoogte of aantal afbeeldingen. Er is ruimte tussen de staven.

Een **histogram** is een bijzonder staafdiagram. Je gebruikt het alleen voor een continue kwantitatieve variabele. De horizontale as is dan een getallenlijn. De staven staan tegen elkaar aan.

Een **lijndiagram** of **frequentiepolygoon** (polygoon = veelhoekig) ontstaat door in een histogram de middens van de bovenkanten van de staven te verbinden met lijnstukken en daarna de staven te verwijderen.

Het **steelbladdiagram** is een variant op de frequentietabel. Eigenlijk is het een frequentietabel en een histogram tegelijk, waarbij de afzonderlijke waarnemingen zichtbaar blijven.

Een **cirkeldiagram** laat relatieve frequenties zien als **sectorhoek**. De sectorhoek bereken je door de relatieve frequentie te vermenigvuldigen met $360°$.

Een **cumulatieve frequentietabel** ontstaat wanneer frequenties eerst worden opgeteld. Een **cumulatief histogram** en een **cumulatief frequentiepolygoon** ontstaan door van een cumulatieve frequentietabel een histogram en frequentiepolygoon te maken. Bij het maken van de cumulatieve frequentiepolygoon verbind je de rechterbovenkanten van de staven.

Bij een diagram van een continue variabele en bij een diagram met een klassenindeling zet je de klassengrenzen links en rechts van het punt of de staaf. Anders (dus bij een discrete variabele zonder klassenindeling) staan de waarnemingsgetallen midden onder de punten of staven.

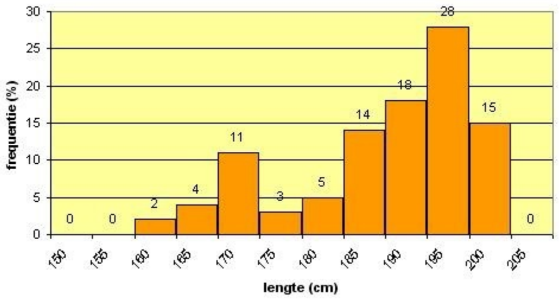
De grafische rekenmachine kent diverse statistische functies en diagrammen, maar bij grote datasets kun je beter werken met computerprogramma's zoals Excel. Zie het **Practicum**.

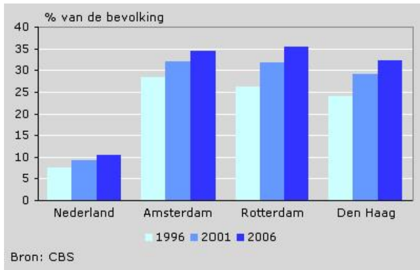
Het Engelse woord voor diagram is ‘chart’ of een ‘graph’. De bijbehorende diagrammen zijn dan bar graph, line graph en pie graph of pie chart.

Voorbeeld 1

Bekijk het diagram waarin de variabele ‘lengte van basketballers’ is weergegeven. De variabele is kwantitatief. Daarom is de variabele in beeld gebracht met een histogram, de staven zijn tegen elkaar aan getekend. De volgorde van de staven ligt vast.

Bekijk ook het diagram met percentages allochtonen per stad in een aantal jaren. Dit is een bijzonder diagram, want er zijn eigenlijk twee variabelen in verwerkt: *het jaar* (kwantitatief) en *de stad* (kwalitatief). Het diagram is dan ook een combinatie van een staafdiagram en een histogram. Je mag de staven van de jaren niet verwisselen. Je kunt de steden onderling wel verwisselen.





Figuur 3.7

Waarom mag je in het staafdiagram de staven onderling verwisselen en in het histogram niet?

Antwoord

De variabelen in het staafdiagram zijn kwalitatief. Onderling verwisselen van staven bij een kwalitatieve variabele maakt niet uit. Maar de variabele lengte (cm) in het histogram is kwantitatief en loopt op van de kleinste lengte naar de langste lengte. Omdat de variabele op de horizontale as oploopt vanaf 0, kun je de bij de lengte behorende staven niet verwisselen.

Opgave 3

Bekijk **Voorbeeld 1**.

a Wat is het verschil tussen een histogram en een staafdiagram?

b Welke extra afleesmogelijkheid is vaak waardevol bij histogrammen en hun bijbehorende frequentiepolygoon?

c Wat kun je uit een steelbladdiagram wel aflezen, maar uit het bijbehorende histogram niet?

Opgave 4

Dit is een deel van de dienstregeling van een busdienst. Het is een steelbladdiagram.

a Schrijf de frequentie per heel uur in een frequentietabel.

b Maak een histogram met het aantal busritten per uur.

c Welke informatie staat wel in een steelbladdiagram en niet in een histogram?

d Welke informatie staat zowel in een steelbladdiagram als in een histogram?

e Waarom maak je bij elk onderzoek niet altijd een steelbladdiagram in plaats van een histogram?

maandag t/m vrijdag					
05					
06	17	47			
07	17	32[e]	47		
08	02[e]	17	32[e]	47	
09	02[e]	17	28[d]	47	58[d]
10	17	28[d]	47	58[d]	
11	17	28[d]	47	58[d]	
12	17	28[d]	47	58[d]	
13	17	28[d]	43	58[be]	
14	13	28[e]	43	58[be]	
15	13	28[e]	43	58[be]	
16	13	28[e]	43	58[be]	
17	13	28[e]	43	58[e]	
18	13	50			
19	20[ad]	50			
20	20[ad]	50			
21	20[ad]	50			
22	20[ad]	50[c]			
23	20[ad]	50[c]			
00	20[e]	55[d]			
01					

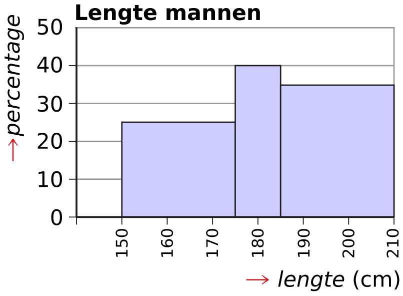
Figuur 3.8

Opgave 5

Bekijk het histogram met de lengtes van een groep mannen. De klassenindeling in dit histogram voldoet niet aan de regels.

a Elk van de drie staven kun je met verticale lijntjes verdelen in staven van 5 cm breed. Leg uit waarom het histogram dan niet meer de juiste weergave van de lengtes van deze groep mannen weergeeft.

b Maak met de gegevens die je uit het gegeven histogram hebt, een correcte versie van dit histogram.



Figuur 3.9

Voorbeeld 2

Maak een histogram en een frequentiepolygoon van de relatieve frequenties bij de variabele *levendgeborenen naar leeftijd moeder*.

Antwoord

Maak eerst de lijst met relatieve frequenties.

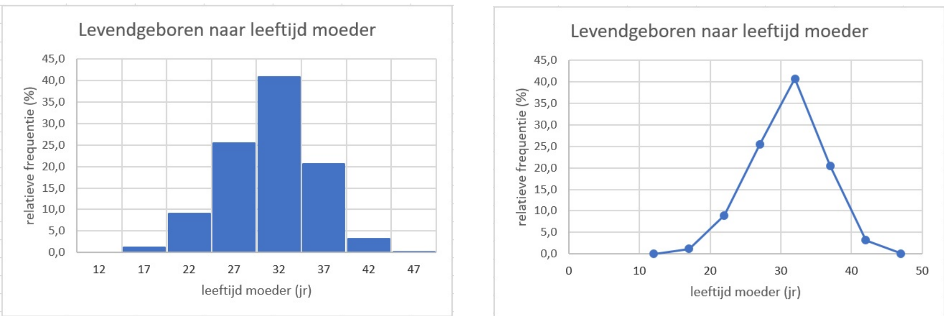
Maak vervolgens het histogram en de frequentiepolygoon. Bekijk het resultaat.

levendgeboren naar leeftijd moeder	2003
klasse	frequentie
tot 15 jaar	0
15 - 19 jaar	2182
20 - 24 jaar	17383
25 - 29 jaar	49344
30 - 34 jaar	79001
35 - 39 jaar	39665
40 - 44 jaar	6225
45 - 49 jaar	217
	194017

Figuur 3.10

klasse		*freq.*	*rel. freq. (%)*
1	15 – 19 jaar	2182	1,1
2	20 – 24 jaar	17383	9,0
3	25 – 29 jaar	49344	25,4
4	30 – 34 jaar	79001	40,7
5	35 – 39 jaar	39665	20,4
6	40 – 44 jaar	6225	3,2
7	45 – 49 jaar	217	0,1
	totaal	194017	100

Tabel 3.1



Figuur 3.11

Merk op dat Excel eigenlijk geen histogram kan tekenen. Je maakt een staafdiagram dat zo veel mogelijk op een echt histogram lijkt. Het verschil is dat op de horizontale as de klassenmiddens staan.

Opgave 6

Bekijk in **Voorbeeld 2** het overzicht van levendgeborenen naar leeftijd van de moeder in 2003. Maak zelf (met de grafische rekenmachine of met Excel) het histogram en de polygoon van de relatieve frequenties bij deze klassenindeling van levendgeborenen naar leeftijd van de moeder.

levendgeboren naar leeftijd moeder	2003
klasse	**frequentie**
tot 15 jaar	0
15 - 19 jaar	2182
20 - 24 jaar	17383
25 - 29 jaar	49344
30 - 34 jaar	79001
35 - 39 jaar	39665
40 - 44 jaar	6225
45 - 49 jaar	217
	194017

Figuur 3.12

Voorbeeld 3

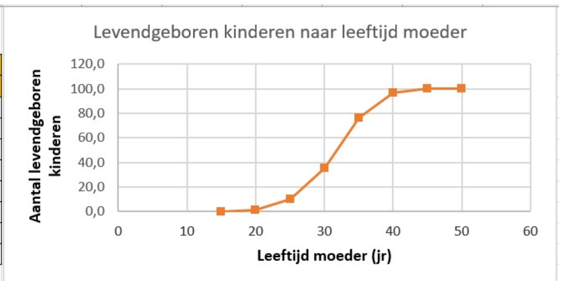
Maak een cumulatief relatief frequentiepolygoon bij de klassenindeling van de variabele *levendgeboren naar leeftijd moeder*.

Antwoord

Maak eerst de lijst met relatieve frequenties en cumulatieve relatieve frequenties. In Excel gaat dat heel gemakkelijk. Maak vervolgens een lijst met rechterklassengrenzen. Nu ga je geen lijndiagram maken, maar je kiest voor spreiding en je maakt een lijngrafiek. Bekijk het resultaat.

(Excel kan eigenlijk geen cumulatief histogram tekenen en ook geen cumulatief frequentiepolygoon. Je maakt een grafiek die zo veel mogelijk op een frequentiepolygoon lijkt.)

levendgeboren naar leeftijd moeder	2003			
		X-as		Y-as
klasse	**frequentie**	**rechtergrens**	**cum.freq.**	**cum.freq.(%)**
tot 15 jaar	0	15	0	0,0
15 - 19 jaar	2182	20	2182	1,1
20 - 24 jaar	17383	25	19565	10,1
25 - 29 jaar	49344	30	68909	35,5
30 - 34 jaar	79001	35	147910	76,2
35 - 39 jaar	39665	40	187575	96,7
40 - 44 jaar	6225	45	193800	99,9
45 - 49 jaar	217	50	194017	100,0
	194017			



Figuur 3.13

Opgave 7

Bekijk **Voorbeeld 3**. Maak het histogram en het cumulatieve relatieve frequentiepolygoon bij deze klassenindeling van levendgeborenen per leeftijd van de moeder (bijvoorbeeld met Excel).

Opgave 8

Bekijk de gegevens over het verbruik van energie in 1998. Het verbruik is gegeven in PJ (1 PJ = 1 petajoule = 10^{15} joule).

1998	*aantal PJ*	*percentage* (%)	*graden* (°)
aardgas	1284	48,8	
aardolie	939		
steenkool	347		
overige	63		
totaal	2633	100	360

Tabel 3.2

a Bereken de procentuele verhoudingen tussen aardgas, aardolie, steenkool en overige in 1998.

b Maak een cirkeldiagram voor het energieverbruik in het jaar 1998.

In de tabel zie je het verbruik in PJ in de loop van de jaren.

kalenderjaar	*aardgas*	*aardolie*	*steenkool*	*overige*
1998	1284	939	347	63
2000	1469	1073	332	105
2001	1508	1113	351	106
2002	1500	1118	355	104
2003	1481	1194	370	112

Tabel 3.3

c Teken een nieuw cirkeldiagram voor de onderlinge verhoudingen tussen aardgas, aardolie, steenkool en overige voor het jaar 2003 met behulp van deze tabel.

d Vergelijk de procentuele verhouding tussen aardgas, aardolie, steenkool en overige van de jaren 1998 en 2003. Wat valt je op?

Verwerken

Opgave 9

Bekijk de opbouw van de benzineprijs van Euro 95 volgens de Bovag.

a Hoeveel procent is de bruto winstmarge voor het tankstation volgens de Bovag?

b Geef de opbouw weer in een cirkeldiagram.

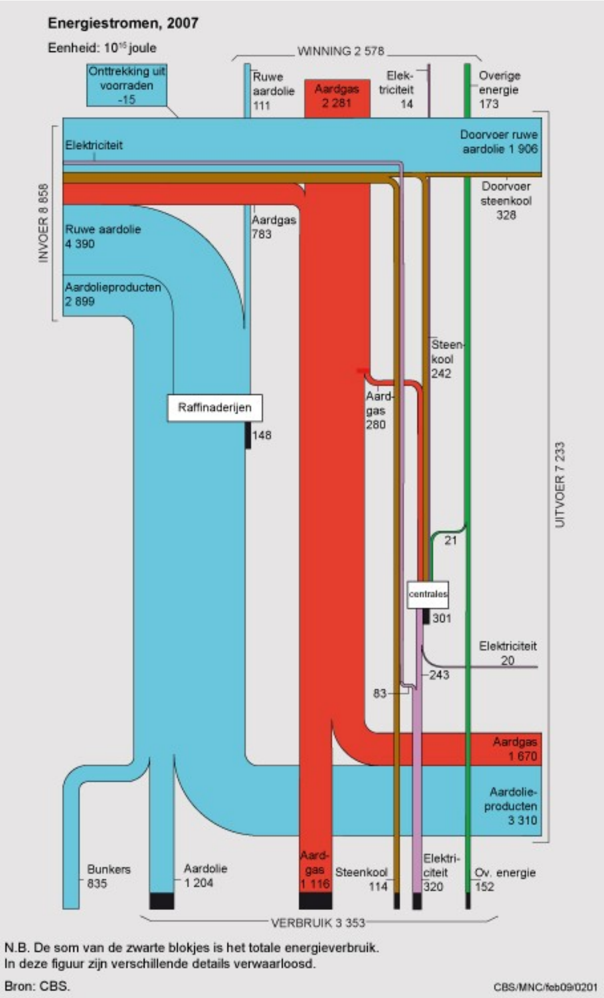
c Wat zal de Bovag zeggen als consumenten klagen over de hoge benzineprijzen?

Opbouw benzineprijs Euro 95	
bedragen in eurocenten	
(Aan deze berekening kunnen geen rechten worden ontleend)	
Adviesprijs	*144*
Productprijs	45,3
Distributiekosten	6,8
Brutowinst oliemaatschappij	1,4
Bruto winstmarge tankstation	5
Accijns en heffingen	62,5
BTW	23

Figuur 3.14

Opgave 10

Dit stroomdiagram geeft de energiebalans van Nederland weer. Je ziet de hoeveelheid energie die Nederland opwekt en invoert. Je ziet ook de energie die we met z'n allen verbruiken of doorvoeren/ uitvoeren naar het buitenland. De gebruikte eenheid is 10^{15} joule.



Figuur 3.15

a Wat betekent het getal 2363 bij de aardgaswinning?

b Hoeveel joule energie is er in 2009 verbruikt door onze energiecentrales om elektriciteit op te wekken?

c Deze energiecentrales halen hun energie behalve uit aardgas en steenkool ook uit andere energiebronnen. Waaruit blijkt dat? En welke energiebronnen zijn dat?

d Hoeveel joule energie is er in Nederland in 2009 verbruikt?

e Hoeveel joule energie is er in Nederland in 2009 ingevoerd?

f Hoeveel joule energie is er als elektriciteit ingevoerd?

g Waarom was het vinden van aardgas in de Nederlandse bodem de afgelopen jaren zo belangrijk voor onze economie?

h Nederland kent ook opgeslagen energievoorraden. Waar zie je dat in het schema?

Opgave 11

Bekijk de frequentieverdeling van de weeklonen van 65 werknemers van een bedrijf.

a Bereken de relatieve frequenties bij deze tabel.

b Maak een staafdiagram van de frequenties en van de relatieve frequenties.

c Maak een frequentiepolygoon.

Het bedrijf neemt vijf extra werknemers in dienst. Zij krijgen een weekloon van € 835,00; € 1156,00; € 1345,00; € 1567,00 en € 1714,00.

d Pas de frequentietabel aan voor de zeventig werknemers.

e Teken een staafdiagram en een lijndiagram bij de nieuwe frequentietabel.

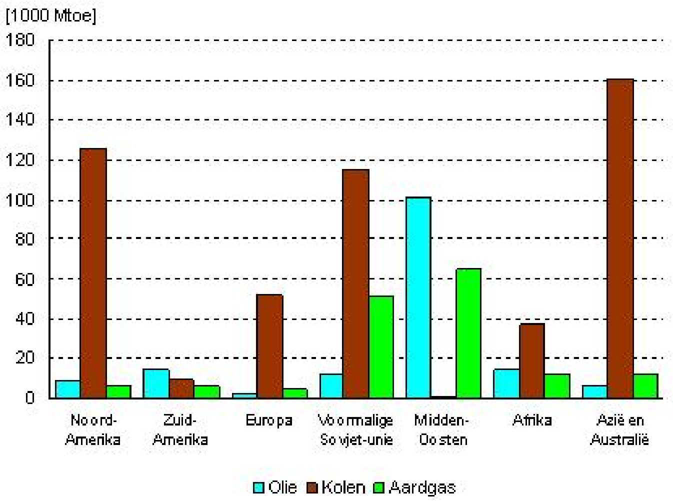
loon (€)	*aantal*
500– < 600	8
600– < 700	10
700– < 800	16
800– < 900	14
900– < 1000	10
1000– < 1100	5
1100– < 1200	2
totaal	65

Tabel 3.4

Opgave 12

Je ziet een staafdiagram van de wereldvoorraad olie, kolen en gas per regio per eind 2003.

Mtoe = miljoen ton olie-equivalenten = 41808 terajoules.



Figuur 3.16

a Hoe zie je dat dit diagram geen histogram is?

b Waarom is een staafdiagram gemaakt en geen cirkeldiagram?

c Je ziet de bij het staafdiagram behorende data voor eind 2003 en eind 2002. Maak een cirkeldiagram voor de aardgasvoorraad per regio per eind 2003.

Wereld olie-, kolen- en gasvoorraden per regio per eind 2003 Referentie: BP review of World Energy, BP						
	2002			2003		
[1000 Mtoe]	Olie	Kolen	Aardgas	Olie	Kolen	Aardgas
Noord-Amerika	6,4	126	6,6	8,9	126	6,6
Zuid-Amerika	14,1	9,8	6,5	14,3	9,8	6,5
Europa	2,7	59,6	5,3	2,3	51,6	5,1
Voormalige Sovjet-unie	10,3	36,8	10,9	12,2	115,2	50,9
Midden-Oosten	93,4	1,1	51,5	101,7	1,1	64,6
Afrika	10,6	107,2	50,7	14,3	36,9	12,4
Azië en Australië	5,2	160,6	11,6	6,7	160,6	12,1
Totaal	142,7	501,2	143	160,4	501,2	158,2

Figuur 3.17

d Welk soort diagram zou je maken als je de voorraden per regio per eind 2003 en per eind 2002 met elkaar wilt vergelijken?

(bron: energie.nl)

Opgave 13

In de tabel zie je de behaalde cijfers voor een wiskundetoets door twee parallelklassen.

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

Tabel 3.5

a Verwerk de resultaten van beide klassen in één frequentietabel waarin de resultaten van beide klassen gescheiden blijven en teken het bijbehorende staafdiagram. Kies een klassenbreedte van $0{,}5$.

b Om een overzicht te krijgen van hoe de toets gemaakt is, kun je de resultaten verwerken in een steelbladdiagram. Doe dat.

c Om het verschil tussen beide klassen te onderzoeken, kun je de resultaten verwerken in een dubbel steelbladdiagram. Doe dat.

d Noem enkele voordelen die het steelbladdiagram heeft boven een frequentietabel en een histogram.

Opgave 14

Je ziet informatie over de bevolking van Amsterdam.

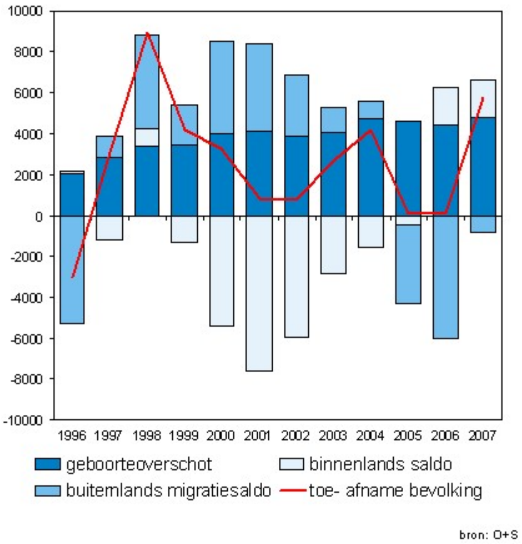
a Welke diagrammen herken je in de figuur?

b Wat betekenen de variabelen *geboorteoverschot*, *buitenlands migratiesaldo* en *binnenlands saldo*?

c Wat is de bevolkingstoename van Amsterdam in 2004 ongeveer? Geef voor dat jaar het *geboorteoverschot*, het *buitenlands migratiesaldo* en het *binnenlands saldo*.

d Het migratiesaldo zit soms boven en soms onder de nullijn. Leg uit wat dat betekent.

e Aan het lijndiagram zie je dat in 2007 de Amsterdamse bevolking met ongeveer 6000 personen is toegenomen. Laat zien hoe je dit kunt berekenen met het staafdiagram.



Figuur 3.18

Toepassen

Opgave 15: Eigen onderzoek

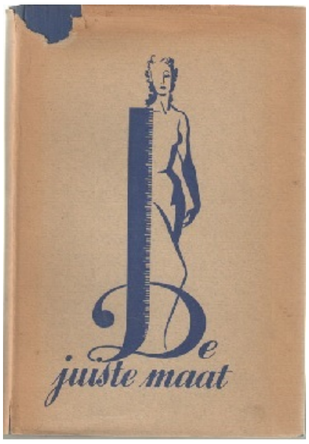
Inmiddels heb je hopelijk zelf een onderzoek uitgevoerd en de resultaten in tabellen weergegeven.

Nu ga je diagrammen maken bij deze tabellen.

1. Bedenk bij elke tabel die je hebt gemaakt wat voor soort diagram de gegevens volgens jou het beste weergeeft. Geef hier argumenten voor. Gebruik hierbij de lijst met variabelen die je eerder hebt opgesteld en waarbij je hebt aangegeven of de variabelen kwalitatief, discreet kwantitatief of continu kwantitatief zijn.
2. Maak de diagrammen die je bij je tabellen hebt gekozen. Dit kan op papier, of met Excel.

Let op! Je hoeft nog geen uitspraken te doen. Bewaar je resultaten wel goed, dan kun je er later mee verder.

Opgave 16: De juiste maat



Figuur 3.19

In 1951 verscheen bij uitgeverij Stafleu in Leiden het boek 'De Juiste Maat', met als ondertitel 'Lichaamsafmetingen van Nederlandse vrouwen als basis voor een nieuw maatsysteem voor damesconfectiekleding'. Auteurs van dit boek waren J. Sittig, Adviesbureau voor Toegepaste Statistiek, en prof. dr. H. Freudenthal, Rijksuniversiteit Utrecht. Het onderzoek was gehouden in opdracht van N.V. Magazijn De Bijenkorf, Amsterdam. In het kader van dit onderzoek zijn bij 5001 vrouwelijke klanten van de Bijenkorf vijftien lichaamsmaten opgemeten. Vervolgens is gekeken welke van deze maten het meest bruikbaar zijn om een **maatsysteem voor kleding** op te baseren.

Bekijk een deel van de uitkomst van het onderzoek in het bestand **Statistiek Bijenkorf 1947**.

a Om welk type statistische variabelen gaat het hier?

b Welke klassenindeling is er gebruikt voor de variabele *lengte*?
En voor de variabele *gewicht*?

c Maak een histogram van de relatieve frequenties van de lengtes van de vrouwen.
Wat valt op aan dit histogram?

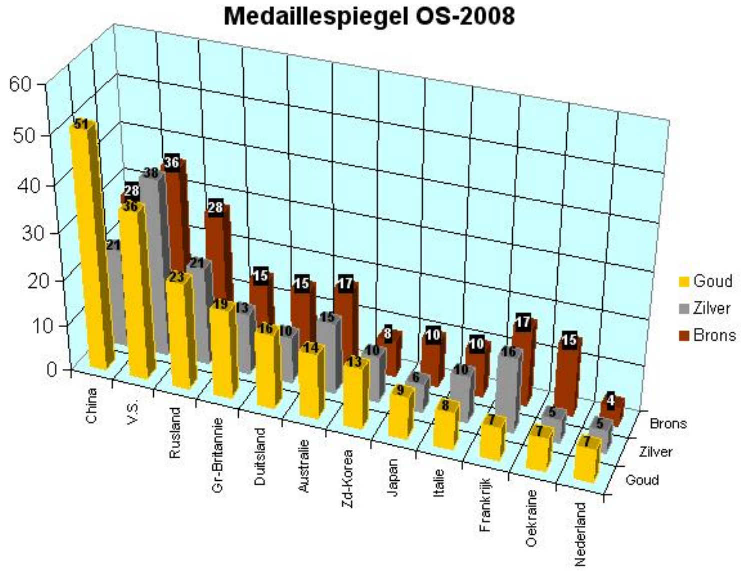
d Hoeveel procent van de vrouwen heeft een lengte vanaf (afgerond) $1{,}56$ m tot en met $1{,}68$ m?

e Maak een histogram van de relatieve frequenties van de gewichten van de vrouwen.
Wat valt op aan dit histogram?

Testen

Opgave 17

Je ziet de medaillespiegel van de Olympische Spelen van 2008 in Beijng met de beste 12 landen.



Figuur 3.20

a Wat geeft elke staaf in dit diagram weer?

b Waarom is een 3D-diagram hier handig? Wat staat er op elk van de assen weergegeven?

c Welk land heeft de meeste gouden medailles gewonnen?

d Welk land heeft de meeste zilveren medailles gewonnen?

e Welk land heeft totaal de meeste medailles gewonnen?

f Deze gegevens kun je ook in een gestapeld staafdiagram weergeven. Hoe ziet dat eruit? Wat is het voordeel en het nadeel?

g Bedenk een presentatie die alle gewenste informatie bevat en een duidelijk overzicht geeft.

Opgave 18

Voor een toets kun je maximaal 100 punten scoren. Je ziet de scores van een groep van veertig personen.
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

a Deel deze scores in klassen in. Neem als laagste klasse $25- < 35$. Maak een frequentietabel.

b Maak bij deze tabel een histogram van relatieve frequenties.

c Maak een frequentiepolygoon met de relatieve frequenties.

d Personen die 55 of meer punten hebben, scoren voldoende. Maak een cumulatief relatief frequentiepolygoon en bepaal hoeveel procent van deze groep voldoende heeft gescoord.

Practicum

Met de volgende practica kun je **de statistische functies van de grafische rekenmachine** doornemen. Onder andere staat er in hoe je gegevens in lijsten invoert en statistische diagrammen maakt.

- **Statistiek en de TI-84**
- **Statistiek en de TInspire**
- **Statistiek en de Casio**
- **Statistiek en de HPprime**
- **Statistiek en de NumWorks**

De **statistische functies van Excel** vind je in het volgende practicum. Er staat meer in dan op dit moment nodig is, maar onder andere kun je er nog eens in vinden hoe je diagrammen maakt.

- **Data presenteren**

2.4 Gegevens samenvatten

Inleiding

Grote verzamelingen statistische gegevens zijn (ondanks het gebruik van tabellen en diagrammen) vaak nogal onoverzichtelijk. Vergelijken van frequentieverdelingen is ook niet altijd in één oogopslag mogelijk. Daarom worden de resultaten van een statistisch onderzoek vaak samengevat in een aantal getallen die snel informatie geven over frequentieverdelingen, zoals gemiddelden en spreiding van gegevens.

Een aantal opgaven uit dit onderdeel is afkomstig uit de NLT-module: ‘Maak het verschil’, net als enkele datasets.

Je leert in dit onderwerp

- een reeks waarnemingen samen te vatten met centrummaten: gemiddelde, mediaan en modus;
- een reeks waarnemingen samen te vatten met spreidingsmaten: variantie, spreidingsbreedte, kwartielafstand en standaardafwijking;
- een reeks waarnemingen samen te vatten in een boxplot.

Voorkennis

- gegevens in een frequentietabel (met zowel absolute als relatieve als cumulatieve frequenties) verwerken;
- een (cumulatief) histogram en een (cumulatieve) frequentiepolygoon maken bij een frequentietabel.

Verkennen

Opgave V1

De schoolleiding en de ouderraad willen een goed beeld krijgen van de aanwezigheid van leerlingen in jullie leerjaar bij de diverse vakken. Ze willen dat in een beknopt en helder verslag van maximaal twee A4'tjes uiteengezet hebben.

Hoe kan zo’n beknopt verslag eruit zien, zodat iedereen goed geïnformeerd wordt? Denk aan het samenvatten van gegevens met behulp van diagrammen, gemiddelden en uitersten.

Uitleg

Jouw klas heeft een toets gehad. Je docent doet de mededeling dat de toets goed is gemaakt met een gemiddelde van 7,3. Ben je blij met deze informatie of hoor je liever dat het modale cijfer 7,3 is? Of dat de mediaan 7,3 is?

Met deze mededeling probeert je docent een frequentieverdeling met één getal te karakteriseren.

- Het modale cijfer is het cijfer dat het vaakst voorkomt. Hier zegt het niet veel, want misschien komt 7,3 twee keer voor en zijn alle andere cijfers heel verschillend.
- Ook een mediaan (middelste cijfer) van 7,3 zegt niet veel, hoewel je dan zeker weet dat de helft van de cijfers hoger dan of gelijk aan 7,3 is (en de andere helft lager dan of gelijk aan 7,3).
- Het gemiddelde krijg je door alle cijfers op te tellen en te delen door het totale aantal leerlingen. Maar ben jij een gemiddelde of bovengemiddelde leerling?

Deze getallen zeggen op zichzelf weinig. Het wordt al beter als je er een mededeling over de spreiding van de cijfers bij krijgt. Het gemiddelde cijfer is een 7,3 en de cijfers hebben een spreiding van 2,2 bijvoorbeeld.

Maar wat wordt onder de spreiding verstaan? Het verschil tussen het hoogste en het laagste cijfer, de spreidingsbreedte, is bijvoorbeeld zo'n spreidingsmaat. Maar er zijn ook andere spreidingsmaten.
Een boxplot, waarin onder andere mediaan en spreiding zijn verwerkt, kan je meer duidelijkheid verschaffen.

Opgave 1

Bekijk de **Uitleg**. In de tabel zie je de cijfers van een wiskundetoets van twee parallelklassen.

cijfers klas A						**cijfers klas B**				
6,7	6,4	4,9	3,8	4,0		4,0	6,2	4,9	3,9	5,9
5,6	5,8	6,8	8,2	4,7		7,3	4,7	6,7	7,6	9,4
3,4	8,5	4,1	6,9	7,3		8,3	5,7	7,2	8,7	7,1
6,1	7,5	6,7	6,2	3,4		7,0	6,5	7,4	5,0	4,8
7,9	4,5	8,3				7,7	6,5	4,9	8,8	6,3

Tabel 4.1

a Waarom heeft het geen zin om van beide klassen het modale cijfer te vergelijken?

b Bepaal van beide klassen de mediaan.

c Zegt de mediaan iets over welke klas beter heeft gescoord?

d Bereken van beide klassen het gemiddelde cijfer.

e Welke van beide klassen heeft het hoogste gemiddelde? Kun je nu zonder meer zeggen dat die klas ook beter heeft gescoord?

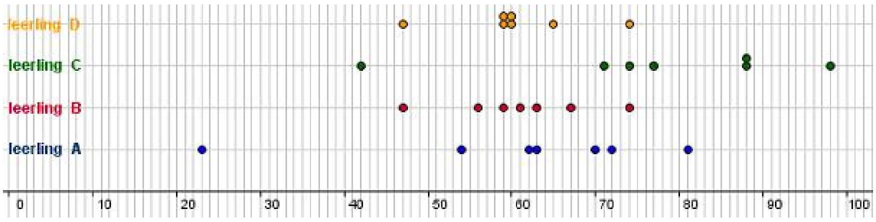
Opgave 2

Je ziet de SE-cijfers (schoolexamen) van enkele leerlingen aan het eind van havo 5. Hun eindcijfer SE is het gemiddelde van deze cijfers.

	se1	se2	se3	se4	se5	se6	se7
A	7,2	6,3	7,0	2,3	6,2	8,1	5,4
B	5,9	7,4	5,6	6,7	6,1	6,3	4,7
C	8,8	9,8	7,4	8,8	5,7	7,3	4,2
D	7,5	6,0	6,0	6,5	6,1	6,1	4,7

Figuur 4.1

Elk SE-cijfer telt even zwaar mee. In de figuur is voor elke leerling elk SE-cijfer aangegeven door een bolletje op een getallenlijn (de komma in het cijfer is weggelaten).



Figuur 4.2

a De leerlingen A en B hebben hetzelfde gemiddelde. Toch is hun cijferbeeld nogal verschillend. Hoe komt dat?

b De spreiding van de cijfers van leerling A en C is vrijwel hetzelfde. Waarin verschilt hun cijferbeeld vooral?

c De cijfers van de leerlingen B en D hebben dezelfde spreidingsbreedte. Is de spreiding van hun cijfers ook hetzelfde?

Een andere maat voor de spreiding vind je door te kijken hoe ver elk cijfer van het gemiddelde af ligt. Bereken van elk cijfer het verschil met het gemiddelde. Je ziet die verschillen voor leerling A.

	se1	se2	se3	se4	se5	se6	se7
A	7,2	6,3	7,0	2,3	6,2	8,1	5,4
	1,1	0,2	0,9	-3,8	0,1	2,0	-0,7

Figuur 4.3

d Bereken het gemiddelde van deze verschillen. Verbaast het antwoord je? Licht je antwoord toe.

Het gemiddelde van deze verschillen is geen goede spreidingsmaat. Dat zit hem in de mintekens. Door te kwadrateren vallen die mintekens weg. Maak voor leerling A een lijst van de kwadraten van de verschillen.

e Bereken daarvan het gemiddelde. Heb je nu een goede spreidingsmaat?

Door het kwadrateren wordt het getal dat je bij e hebt gevonden, nogal groot. Dat los je op door de wortel uit dit getal te nemen. Je krijgt dan de standaardafwijking van de set cijfers.

f Ga na dat de standaardafwijking voor leerling A ongeveer $1{,}73$ is.

g Bereken ook voor leerling B de verschillen van de cijfers met het gemiddelde. Bereken vervolgens het gemiddelde van de kwadraten van die verschillen en de standaardafwijking.

Theorie en voorbeelden

Om te onthouden

Getallen die (bij benadering) het midden aangeven van een reeks waarnemingen heten centrummaten. Er zijn drie **centrummaten**.

- De **modus** is de waarneming met de hoogste frequentie. Vooral geschikt voor kwalitatieve variabelen.
- De **mediaan** is het middelste waarnemingsgetal als de waarnemingsgetallen op volgorde van klein naar groot staan. Is het aantal even, dan zijn er twee middelste waarnemingsgetallen. De mediaan is dan het gemiddelde van die middelste twee.
- Het **gemiddelde** bereken je door alle waarnemingsgetallen op te tellen en te delen door het totale aantal. Als je de waarnemingsgetallen x_1, x_2, x_3, ..., x_n noemt, schrijf je dit als: $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

  Daarin geldt $\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$.

  De Griekse hoofdletter sigma ($\sum$) is het somteken. Bij een frequentietabel vermenigvuldig je elk waarnemingsgetal met de frequentie.

  Het gemiddelde is dan: $\overline{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{n}$.

Bij **klassenindelingen** spreek je van de **modale klasse** en kun je de mediaan het beste opzoeken in een cumulatieve relatieve frequentiepolygoon (de waarde bij 50% schatten door aflezen). Het gemiddelde kun je dan alleen maar schatten door het gemiddelde van de **klassenmiddens** te berekenen.

Centrummaten alleen zeggen nog weinig, er hoort steeds een spreidingsmaat bij. Er zijn drie **spreidingsmaten**:

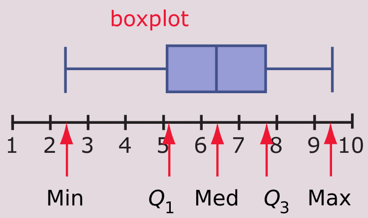
- De **spreidingsbreedte** (ook **variatiebreedte**) is het verschil tussen het hoogste en laagste waarnemingsgetal.
- De **interkwartielafstand** is het verschil tussen de mediaan van de grootste helft (het **derde kwartiel** of Q_3) en de mediaan van de kleinste helft (het **eerste kwartiel** of Q_1). Om de kwartielen te bepalen, zet je eerst de waarnemingsgetallen in volgorde van klein naar groot en verdeel je ze in twee helften. Bestaan de waarnemingen uit een oneven aantal waarden, dan wordt de mediaan van de hele set niet meegenomen om Q_1 en Q_3 te berekenen.
- De **standaardafwijking** (of **standaarddeviatie**) vind je door van elk waarnemingsgetal het verschil met het gemiddelde te bepalen en dat getal te kwadrateren. Die kwadraten tel je op en je deelt ze door het totale aantal waarnemingen. Dit getal heet de **variantie**. De wortel uit de variantie is de standaarddeviatie $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot f_i}{n}}$. De Griekse (kleine) letter sigma is het teken voor standaardafwijking.

Bij klassenindelingen is de spreidingsbreedte het aantal klassen maal de klassenbreedte. De mediaan en de kwartielen zoek je het beste op in een cumulatieve relatieve frequentiepolygoon (de mediaan bij 50%, het eerste kwartiel bij 25% en het derde kwartiel bij 75%). De standaarddeviatie kun je nu alleen schatten door de standaarddeviatie van de klassenmiddens te berekenen. De mediaan, het gemiddelde en alle spreidingsmaten kunnen alleen gebruikt worden voor kwantitatieve variabelen. Hoe je ze met de grafische rekenmachine bepaalt, zie je in het **Practicum**. Het is bij grotere datasets verstandiger om met Excel te werken.

De mediaan, het eerste en derde kwartiel en de spreidingsbreedte en de kwartielafstand kun je laten zien in een **boxplot**. Een boxplot heeft dus vijf grenzen:

- Linkergrens met het laagste getal.
- Rechtergrens met het hoogste getal.
- Middelste grens Q_2, de mediaan.
- De tweede grens Q_1 tussen de linkergrens en Q_2; de mediaan van de eerste helft.
- De vierde grens Q_3 tussen Q_2 en de rechtergrens; de mediaan van de tweede helft.

De interkwartielafstand is het verschil tussen het eerste kwartiel (Q_1) en het derde kwartiel (Q_3), dus $Q_3 - Q_1$.



Figuur 4.4

Voorbeeld 1

Bekijk het steelbladdiagram van de cijfers in een klas. Het is tegelijk een klassenindeling (eerste klasse $2{,}0- < 3{,}0$) en een overzicht van alle cijfers (in de tweede klasse zit twee keer het cijfer $3{,}9$). Je kunt er ook een boxplot van maken. Laat dat zien.

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

Figuur 4.5

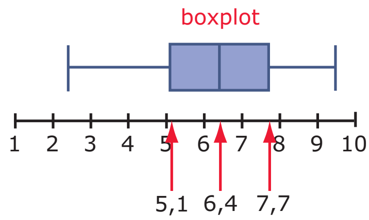
Antwoord

Bepaal eerst:

- De modus is $8{,}6$ en de modale klasse is $6{,}0- < 7{,}0$.
- De mediaan is $6{,}4$ (het gemiddelde van de twee middelste cijfers), het eerste kwartiel is $5{,}1$ en het derde kwartiel is $7{,}7$.
- De spreidingsbreedte is $9{,}5 - 2{,}4 = 7{,}1$ als je naar de werkelijke cijfers kijkt, of $10{,}0 - 2{,}0 = 8{,}0$ als je naar de klassenindeling kijkt.
- De kwartielafstand is $7{,}7 - 5{,}1 = 2{,}6$. Het is de breedte van de box van de boxplot.

De boxplot verdeelt alle cijfers in vier delen met elk 25% van deze cijfers.

Bestaan de waarnemingen uit een oneven aantal waarden, dan wordt de mediaan van de hele set niet meegenomen om Q_1 en Q_3 te berekenen.



Figuur 4.6

Opgave 3

Bekijk in **Voorbeeld 1** het steelbladdiagram en de boxplot. In de tabel zie je de cijfers gehaald voor een wiskundetoets van twee parallelklassen.

cijfers klas A						**cijfers klas B**				
6,7	6,4	4,9	3,8	4,0		4,0	6,2	4,9	3,9	5,9
5,6	5,8	6,8	8,2	4,7		7,3	4,7	6,7	7,6	9,4
3,4	8,5	4,1	6,9	7,3		8,3	5,7	7,2	8,7	7,1
6,1	7,5	6,7	6,2	3,4		7,0	6,5	7,4	5,0	4,8
7,9	4,5	8,3				7,7	6,5	4,9	8,8	6,3

Tabel 4.2

Maak van de cijfers van beide klassen een steelbladdiagram en bepaal de mediaan en de kwartielafstand van beide klassen. Teken voor beide klassen een boxplot van de resultaten.

Opgave 4

Welke uitspraak is waar voor de volgende waarnemingsgetallen?
58; 63; 51; 56; 86; 69; 55; 76; 74; 69; 45; 75; 55; 68; 68; 52; 70; 57; 65; 78; 65; 72; 83; 65; 79.

A. De modus en mediaan zijn gelijk.
B. De modus en het gemiddelde zijn gelijk.
C. Het gemiddelde en de mediaan zijn gelijk.
D. Geen van deze uitspraken is waar.

Opgave 5

Welke uitspraken zijn waar voor de volgende waarnemingsgetallen?
58; 63; 51; 56; 86; 69; 55; 76; 74; 69; 45; 75; 55; 68; 68; 52; 70; 57; 65; 78; 65; 72; 83; 65;
79; 57; 63; 63; 72; 63.

A. De modus is groter dan de mediaan.
B. Het gemiddelde is groter dan de mediaan.
C. De modus is kleiner dan het gemiddelde.

Voorbeeld 2

Bekijk de tabel met leeftijd, lengte en gewicht van 36 mannen.

gegevens van 36 mannen											
nr	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)	*nr*	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)	*nr*	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)
i	L	l	G	i	L	l	G	i	L	l	G
1	20	180	95	13	46	180	95	25	67	171	68
2	23	184,5	91	14	47	177,5	75	26	67	170	81
3	23	180,5	90	15	51	196	86	27	68	169	97
4	29	169	91	16	52	188,5	93	28	68	176	72
5	30	176	80	17	53	170	85	29	71	171	84
6	33	182	82	18	54	178,5	77	30	71	180,5	75
7	36	182,5	85	19	56	172,5	74	31	71	172	77
8	36	176,5	74	20	61	164	85	32	74	171	77
9	38	180	73	21	61	181	89	33	75	183	95
10	40	190,5	112	22	61	173	95	34	75	172,5	74
11	41	184	89	23	61	181	89	35	82	173	66
12	44	179	75	24	63	181,5	100	36	84	175	90

Tabel 4.3

Je wilt van alle drie de series waarnemingsgetallen zowel de drie centrummaten als de drie spreidingsmaten bepalen.

Gebruik je Excel, dan gebruik je de statistische functies om de centrummaten en de spreidingsmaten te (laten) berekenen. Zie het practicum **Practicum**.

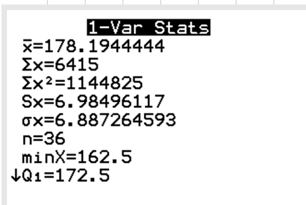
Werk je met de hand of met de grafische rekenmachine, dan maak je meteen geschikte frequentieverdelingen. Je kunt dan de centrummaten en de spreidingsmaten alleen nog schatten vanuit de klassenmiddens. Zie het **Practicum**.

Antwoord

Voor de modus en de mediaan zijn de sorteerfuncties in Excel erg handig. Verder kun je gemakkelijk optellen en kolommen maken met de waarden van een waarnemingsgetal maal zijn frequentie, enzovoort.

In deze **tweede versie van het Excel bestand** vind je de antwoorden. Let op de formules die gebruikt zijn.

Werk je met de grafische rekenmachine, maak dan eerst een tabel van relatieve frequenties met bijbehorende klassenmiddens. Laat vervolgens de machine de centrum- en spreidingsmaten bepalen. Zo krijg je voor de lengte deze gegevens.



Figuur 4.7

Bij de klassenindelingen zijn centrum- en de spreidingsmaten niet exact bepaald, omdat je door indelen in klassen de echte gegevens kwijt bent geraakt. Je krijgt zo schattingen van deze waarden.

Opgave 6

Bekijk in **Voorbeeld 2** de tabel met gegevens van 36 mannen en in het antwoord de tabel met centrummaten en spreidingsmaten.

a Hoe wordt het gemiddelde berekend?

b Hoe wordt de spreidingsbreedte (variatiebreedte) berekend?

c Hoe wordt de kwartielafstand berekend?

d Ga na dat de modale leeftijd, de modale lengte en het modale gewicht correct zijn.

e Bereken de kwartielen en teken de bijpassende boxplots.

f Hoe wordt de standaardafwijking berekend?

Opgave 7

Bekijk de tabel met gegevens van 36 vrouwen.

gegevens van 36 vrouwen											
nr	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)	nr	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)	nr	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)
i	L	l	G	i	L	l	G	i	L	l	G
1	21	181	85	13	53	168,5	64	25	70	166	65
2	22	171	55	14	53	165	64	26	71	159	93
3	30	176	55	15	53	179	77	27	73	154	76
4	32	176,5	70	16	57	174	76	28	73	153	54
5	34	158,5	61	17	62	179	80	29	74	160	60
6	43	160	80	18	64	163,5	80	30	75	163,5	79
7	49	179	75	19	65	160	92	31	75	153	93
8	50	166	74	20	65	182	78	32	76	159	81
9	50	171	84	21	65	154	74	33	77	165	68
10	50	166,5	65	22	67	162	64	34	78	163	89
11	51	171	80	23	68	160	63	35	80	166,5	60
12	52	171	71	24	69	163	75,5	36	80	157	72

Tabel 4.4

a Bepaal vanuit de basisgegevens de centrummaten en de kwartielen van de leeftijden, de lengtes en de gewichten van vrouwen.

b Bepaal ook de bijbehorende spreidingsmaten en teken de boxplots.

Opgave 8

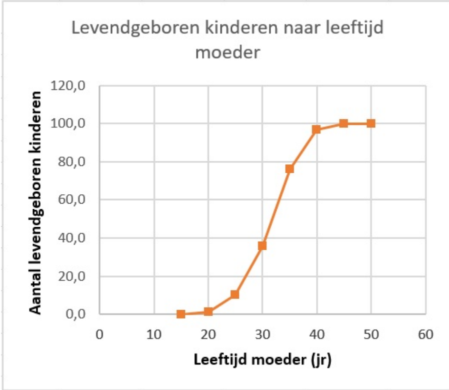
In twee verzorgingstehuizen heeft men de keuze uit verschillende zitbreedtes voor rolstoelen. Van alle zitbreedtes zijn evenveel rolstoelen beschikbaar.
Verzorgingstehuis Alfa heeft rolstoelen met zitbreedtes van 35 cm, 38 cm, 41 cm, 45 cm en 50 cm.
Verzorgingstehuis Omega heeft rolstoelen met zitbreedtes van 34 cm, 38 cm, 41 cm, 46 cm en 50 cm.

a Bereken per verzorgingstehuis het gemiddelde en de mediaan van de zitbreedtes van de rolstoelen.

b Geven deze centrummaten het verschil tussen de zitbreedtes van de rolstoelen van beide verzorgingstehuizen voldoende aan? Licht je antwoord toe.

Voorbeeld 3

Je ziet een cumulatieve relatieve frequentiepolygoon bij deze klassenindeling van levendgeborenen naar leeftijd van de moeder. Maak een bijpassende boxplot.



Figuur 4.8

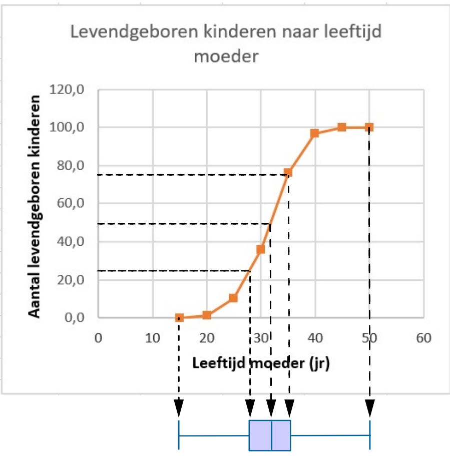
levendgeborenen naar leeftijd moeder 2003	
klasse	*frequentie*
15 – 19 jaar	2182
20 – 24 jaar	17383
25 – 29 jaar	49344
30 – 34 jaar	79001
35 – 39 jaar	39665
40 – 44 jaar	6225
45 – 49 jaar	217

Tabel 4.5

Antwoord

Lees bij 50% de mediaan af, bij 25% het eerste kwartiel en bij 75% het derde kwartiel. Het minimum en maximum zitten bij 0% en 100%. Je ziet:

- het eerste kwartiel $Q_1 = 28$
- de mediaan $Q_2 = 31{,}5$
- het derde kwartiel $Q_3 = 35$



Figuur 4.9

Opgave 9

Bestudeer **Voorbeeld 3** hoe je bij een klassenindeling een boxplot maakt.

Bekijk vervolgens in **Voorbeeld 2** de tabel met gegevens van 36 mannen. Maak bij de klassenindeling van de kolom ‘lengte’ een cumulatieve relatieve frequentiepolygoon en een boxplot.

Verwerken

Opgave 10

Je hebt de waarnemingsgetallen 16, 18, 22, 24, 26, 26, 28, 30 en 36.

a Teken een boxplot.

b Doe dat nog eens als je bij alle getallen 4 optelt.

c En ook als je van alle getallen 40 aftrekt.

d Doe het nog eens als je alle getallen door 2 deelt.

e Welk resultaat krijg je als je alle getallen met 3 vermenigvuldigt?

f Beschrijf wat er met de boxplot gebeurt als bij alle waarnemingsgetallen een getal wordt opgeteld of ervan afgetrokken wordt.

g Beschrijf wat er met de boxplot gebeurt als alle waarnemingsgetallen met een getal worden vermenigvuldigd of door een getal worden gedeeld.

Opgave 11

Voor een practicum biologie worden regenwormen gevangen. De lengte van die regenwormen vind je in de tabel.

a Kijk naar de manier waarop de klassen zijn gemaakt. Hoe nauwkeurig zijn de regenwormen gemeten? Bij welke klasse hoort een regenworm die 3,0 cm lang is?

b Welke klasse is de modale klasse?

c Teken een histogram van de cumulatieve relatieve frequenties. Teken in dezelfde figuur de cumulatieve frequentiepolygoon.

d In welke klasse zit de mediaan? Kun je precies zeggen hoe groot die mediaan is? Schat de mediaan met behulp van de cumulatieve frequentiepolygoon.

e Bereken het gemiddelde en de standaardafwijking.

lengte regenworm (cm)	*aantal*
0,0– < 3,0	4
3,0– < 6,0	8
6,0– < 9,0	17
9,0– < 12,0	22
12,0– < 15,0	23
15,0– < 18,0	17
18,0– < 21,0	6
21,0– < 24,0	2
24,0– < 27,0	1

Tabel 4.6

Opgave 12

Een supermarkt laat onderzoek verrichten naar de besteding per klant en naar de hoeveelheid tijd die een klant aan de kassa nodig heeft om af te rekenen. Er worden op verschillende tijdstippen tellingen gehouden. Je ziet de resultaten.

kassatijd (min)	*aantal klanten*
0– < 1	22
1– < 2	75
2– < 3	58
3– < 4	25
4– < 5	16
5– < 6	4

kassabon (€)	*aantal klanten*
0– < 50	24
50– < 100	61
100– < 150	75
150– < 200	25
200– < 250	12
250– < 300	2
300– < 350	1

Tabel 4.7

a Bepaal bij beide tabellen de modus, de mediaan, het eerste en het derde kwartiel en het gemiddelde.

b Hoe groot is de standaardafwijking bij beide verdelingen?

c Teken bij beide tabellen een boxplot.

De supermarkt heeft een weekomzet van € 150000,00. Een caissière mag 38 uur per week werken.

d Hoeveel caissières moet de supermarkt in dienst nemen als er vanwege de wisselende winkeldrukte een overcapaciteit van 25% wordt aangehouden?

Opgave 13

Op elk uur van een dag is de temperatuur bepaald. De uren van middernacht tot 12 uur 's middags worden aangegeven met am (het Latijnse ‘ante meridiem’ (am) betekent ‘voor het middaguur’), en de uren van 12 uur 's middags tot middernacht met pm (‘post meridiem’ (pm)).

uur	1	2	3	4	5	6	7	8	9	10	11	12
am	16,1	15,8	15,4	15,1	14,8	15,0	16,1	17,4	18,5	19,4	20,3	21,1
pm	21,9	22,6	22,7	22,5	21,9	21,0	19,8	18,8	18,1	17,5	17,0	16,6

Tabel 4.8

a Verwerk deze gegevens in een dubbel steelbladdiagram.

b Maak boxplots van elk dagdeel afzonderlijk en van de totale dag.

c Bereken voor beide dagdelen afzonderlijk het gemiddelde en de standaardafwijking.

d Bereken het gemiddelde en de standaardafwijking van alle metingen van die dag.

e Geef een verklaring voor de verschillen die je vindt. (Dit is bivariate statistiek. Je bekijkt twee variabelen en hun eventuele samenhang/verschil.)

Opgave 14

In een nieuw te bouwen ziekenhuis moeten bedden worden aangeschaft. De facilitaire dienst vraagt zich af welke lengte de bedden moeten krijgen. Hoe langer de bedden, hoe hoger de kosten. In het oude ziekenhuis hebben ze het laatste jaar van 278 patiënten gegevens verzameld. Je vindt ze in het bestand **Patiëntengegevens**.

a Bereken de gemiddelde lengte van de patiënten. Bereken ook de gemiddelde lengte van de vrouwelijke en de mannelijke patiënten apart.

b Men kan natuurlijk alle bedden zo lang maken als de langste patiënt. Hoe lang worden de bedden dan? Noem een bezwaar tegen dit idee.

Handiger is misschien de lengte van het bed zo te kiezen dat 50% van de patiënten erin past. Voor langere patiënten neem je dan een bed met een lengte van de langste patiënt.

c Hoe lang moet het bed dan worden als 50% van de patiënten erin past?

Het hoofd van de facilitaire dienst denkt dat het goedkoper is om verschil te maken in mannen- en vrouwenbedden, wat de lengte betreft.

d Als we hiervan uitgaan en de voorwaarde na b, hoe lang worden dan een ‘mannenbed’ en een ‘vrouwenbed’?

Opgave 15

Het gemiddelde van de getallen 23, 6, 15, 31, 7 en c is gelijk aan 16. Bereken c.

Opgave 16

In een eetcafé komen op een dag tachtig mensen eten. Ze besteden gemiddeld € 22,00. Studenten betalen € 15,00 voor het studentenmenu, dagjesmensen betalen € 25,00 voor het toeristenmenu.
Hoeveel studenten waren er die dag?

Toepassen

In 1947 zijn bij 5001 vrouwelijke klanten van de Bijenkorf vijftien lichaamsmaten opgemeten. Vervolgens is gekeken welke van deze maten het meest bruikbaar zijn om een **maatsysteem voor kleding** op te baseren. Bekijk een deel van de uitkomst van het onderzoek in het bestand **Statistiek Bijenkorf 1947**.

Je kunt met Excel centrum en spreiding van de gegevens berekenen en boxplots maken.

Soms is er sprake van een **uitschieter**, een waarde die wel erg veel afwijkt.

In een boxplot is een uitschieter een waarde die meer dan $1{,}5$ keer de interkwartielafstand onder het eerste kwartiel of boven het derde kwartiel zit.

Opgave 17

Bekijk bij **Toepassen** de gegevens over lengte en gewicht van de 5001 gemeten vrouwen.

a Bepaal de modale lengte en bereken de gemiddelde lengte.

b Bepaal de mediaan en de kwartielen van de lengte en teken een bijpassende boxplot.

c Zijn er bij deze lengtes uitschieters? Licht je antwoord toe.

d Verdeel de lengtes in klassen van 5 cm, te beginnen bij $135- < 140$. Maak bij de nieuwe frequentieverdeling een relatieve cumulatieve frequentiepolygoon. Welke vorm heeft de polygoon?

e Bepaal opnieuw de gemiddelde lengte, de mediaan en de kwartielen. Wijken de resultaten veel af van de antwoorden bij de a en b?

Opgave 18

In de dataset in **Toepassen** vind je ook gegevens over de *mouwlengte* en de *kniehoogte* van de 5001 vrouwen.

a Bepaal modus, mediaan en het gemiddelde van de mouwlengtes m_i.

b Bepaal de spreidingsbreedte, de interkwartielafstand en de standaardafwijking van de mouwlengtes m_i.

c Er is één nogal afwijkende mouwlengte. Is hier sprake van een uitschieter?

Testen

Opgave 19

Op een feestje zijn acht personen aanwezig. Je ziet een tabel met gegevens over de feestgangers.

naam	leeftijd in jaar	lengte in cm	favoriete drankje	gewicht in kg	zakgeld in € per week	vervoer-middel
Jan	15	187	cola	71	€ 15	fiets
Leo	16	178	colal	69	€ 20	fiets
Elske	15	172	sinas	66	€ 17	bus
Daphne	17	171	cola	67	€ 20	scooter
Mart	19	195	cassis	76	€ 19	auto
Durk	57	180	bier	74	€ 50	auto
Henk	17	187	cola	75	€ 18	fiets
Leo	15	179	cola	70	€ 18	brommer

Figuur 4.10

a Welke centrummaat zou je gebruiken om voor elke kolom de feestgangers te typeren?

b Welke spreidingsmaat zou je zo mogelijk gebruiken om voor elke kolom de feestgangers te typeren?

c Hoe zou je de doorsneefeestganger omschrijven?

Opgave 20

Een groep leerlingen wordt tijdens de zwemles gevraagd om zo lang mogelijk hun adem in te houden onder water. Je ziet hier de bijbehorende tijden (in seconden).

$17; 39; 17; 21; 26; 21; 31; 17; 37; 43$
$36; 17; 15; 29; 21; 31; 35; 23; 18; 17$
$26; 21; 28; 23; 22; 16; 37; 33; 27$

a Bepaal de drie centrummaten van deze gegevens.

b Bepaal de spreidingsbreedte en de standaardafwijking.

c Teken een boxplot bij deze gegevens.

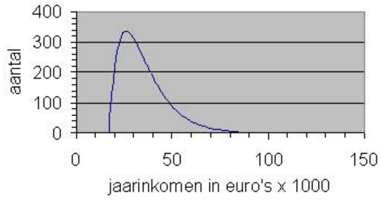
d Maak een klassenindeling met de eerste klasse $15- < 20$.

e Bereken het gemiddelde en de standaarddeviatie bij deze klassenindeling.

f Teken een bijpassende cumulatieve relatieve frequentiepolygoon. Bepaal daarmee de mediaan bij deze klassenindeling.

Opgave 21

In de grafiek vind je de jaarinkomens van de werknemers van een grote fabriek.



Figuur 4.11

a Wat verdient de doorsneewerknemer van deze fabriek? Welke centrummaat heb je gekozen en waarom?

b Welke centrummaat is groter, de mediaan of de modus? Leg uit waar je dat aan kunt zien.

Practicum

Met de volgende practica kun je **de statistische functies van de grafische rekenmachine** doornemen. Onder andere staat er in hoe je centrummaten en spreidingsmaten kunt berekenen.

- **Statistiek en de TI-84**
- **Statistiek en de TInspire**
- **Statistiek en de Casio**
- **Statistiek en de HPprime**
- **Statistiek en de NumWorks**

De **statistische functies van Excel** vind je in het volgende practicum. Er staat meer in dan op dit moment nodig is, maar onder andere kun je er nog eens in vinden hoe je diagrammen maakt en centrummaten en spreidingsmaten kunt laten berekenen door Excel.

- **Data presenteren**

2.5 Uitspraken doen

Inleiding

Een statistisch onderzoek is opgezet om uitspraken te kunnen doen. In kranten en tijdschriften staat het er bol van. Maar vaak ontbreekt belangrijke informatie: er staat bijvoorbeeld wel een gemiddelde, maar er wordt geen spreidingsmaat vermeld. Of er wordt niet vermeld hoe de steekproef is samengesteld.
Welke uitspraken kun je wel doen en welke niet? En wat moet je allemaal vermelden om de betrouwbaarheid van een uitspraak duidelijk te maken?

Je leert in dit onderwerp

- wat een klokvormige frequentieverdeling is en welke vuistregels daarbij horen;
- hoe je verantwoorde uitspraken kunt doen bij statistische gegevens;
- kritisch te kijken naar uitspraken en conclusies die je tegenkomt.

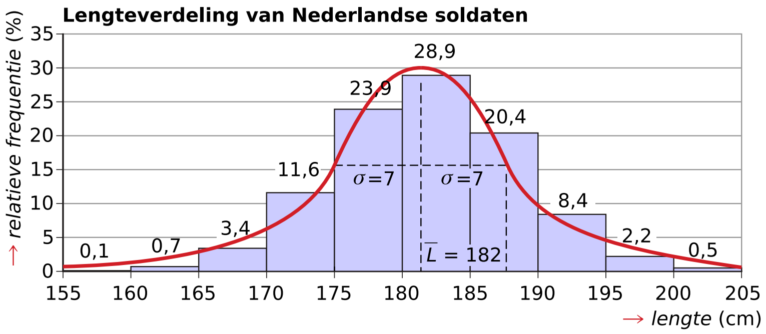
Voorkennis

- allerlei statistische diagrammen interpreteren en maken;
- centrummaten en spreidingsmaten berekenen.

Verkennen

Opgave V1

Bekijk de lengteverdeling van een grote groep mannelijke Nederlandse soldaten.

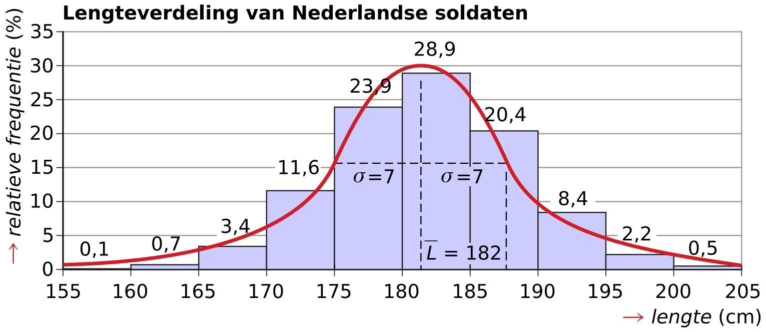


Figuur 5.1

Kun je verklaren waarom de lengtes van zo'n grote groep mannen de vorm van het silhouet van een kerkklok heeft? Waar zit het gemiddelde?

Uitleg

Bekijk de lengteverdeling van een grote groep mannelijke Nederlandse soldaten.



Figuur 5.2

Het histogram of de bijbehorende frequentiepolygoon heeft een klokvorm. Bij veel continue variabelen, bijvoorbeeld bij gewicht, lengte of inhoud, krijg je zo'n klokvormige frequentieverdeling. Je kunt met behulp van gemiddelde en standaardafwijking twee algemene uitspraken doen. Deze uitspraken zijn vuistregels.

Vuistregel 1: tussen $\overline{x}-\sigma_x$ en $\overline{x}+\sigma_x$ zit 68% van de waarnemingsgetallen.

Vuistregel 2: tussen $\overline{x}-2\sigma_x$ en $\overline{x}+2\sigma_x$ zit 95% van de waarnemingsgetallen.

Deze uitspraken betreffen alleen de steekproef (beschrijvende statistiek). 68% van deze soldaten heeft een lengte tussen 175 cm en 189 cm. 95% van deze soldaten heeft een lengte tussen 168 cm en 196 cm.

Vaak wordt verondersteld dat deze gegevens voor de hele populatie Nederlandse jonge mannen gelden en dat die uitspraken op hen van toepassing zijn. Uitspraken doen over een populatie op grond van een steekproef kan alleen als die steekproef representatief is. En dat is hier nog maar de vraag.

Statistiek lijkt spijkerhard, maar je kunt sneller misleid worden door diagrammen en cijfers dan je denkt. Soms wordt een deel van een diagram of van een as weggelaten. Of de cijfers en uitspraken gaan over een te kleine of verkeerde steekproef. Wees altijd op je hoede met cijfers en diagrammen bij een onderzoek. Je hoort immers zelden dat uit een onderzoek geen conclusies getrokken kunnen of mogen worden.

Opgave 1

Gebruik de lengteverdeling van 90 meisjes die je vindt in het bestand **Lengteverdeling 90 meisjes**.

a Maak van deze verdeling een frequentiepolygoon.

De frequentieverdeling is niet een perfecte klokvorm, omdat de steekproef veel te klein is.

b Bereken het gemiddelde en de standaardafwijking van de gegeven lengteverdeling.

c Onderzoek of voor de gegeven lengteverdeling de 68%-vuistregel geldt.

d Onderzoek of voor de gegeven lengteverdeling de 95%-vuistregel geldt.

Opgave 2

Op grond van een representatieve steekproef uit alle Nederlandse meisjes heeft een onderzoeksbureau geconcludeerd dat hun lengtes klokvormig verdeeld zijn met een gemiddelde lengte van 172 cm en een standaardafwijking van 6 cm.

a Bepaal met behulp van de vuistregels hoeveel procent van de Nederlandse meisjes langer is dan $172 + 6 = 178$ cm.

b Bepaal hoeveel procent korter is dan $172 - 2 \cdot 6 = 160$ cm.

Opgave 3

Je ziet hier een aantal conclusies uit de statistische gegevens. Geef commentaar op de uitspraken.

a In 1971 nam de NAVO 49% van alle militaire uitgaven voor haar rekening. In 1981 was dat nog 43%. De militaire uitgaven van de NAVO zijn in 1981 lager dan in 1971.

b Van alle verkeersongelukken op deze weg blijkt bij 25% alcohol een rol te hebben gespeeld. Rijden na het drinken van alcohol is veiliger dan rijden zonder alcohol.

c Wasmiddel XXX wast 20% witter dan alle andere wasmiddelen.

d School A heeft hogere percentages geslaagden dan school B. Je kunt beter op school A zitten als je snel wilt slagen.

Theorie en voorbeelden

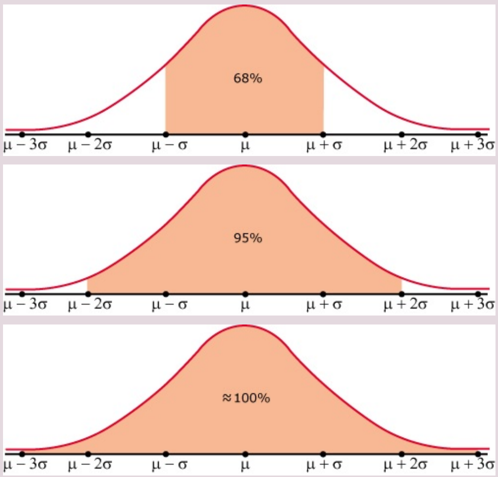
Om te onthouden

In elk deel van een boxplot zit 25% van de waarnemingen. Wanneer het histogram of de bijbehorende frequentiepolygoon bij benadering **klokvormig** is, zijn het gemiddelde $\overline{x}$ en de standaardafwijking σ_x goede karakteristieken van de frequentieverdeling. Er gelden vuistregels.

- Vuistregel 1: tussen $\overline{x} - \sigma_x$ en $\overline{x} + \sigma_x$ zit 68% van de waarnemingsgetallen.
- Vuistregel 2: tussen $\overline{x} - 2\sigma_x$ en $\overline{x} + 2\sigma_x$ zit 95% van de waarnemingsgetallen.
- Vuistregel 3: tussen $\overline{x} - 3\sigma_x$ en $\overline{x} + 3\sigma_x$ zit bijna 100% van de waarnemingsgetallen.

Deze uitspraken betreffen de steekproef.

De uitspraken die je doet over je steekproef zijn alleen geldig voor de hele populatie als de steekproef een goede afspiegeling van die populatie is, dus **representatief** is. De uitspraken zijn betrouwbaarder als de steekproef voldoende groot is.



Figuur 5.3

Voorbeeld 1

De lengteverdeling van Nederlandse mannen boven de 20 jaar is bij benadering klokvormig. De gemiddelde lengte is 180,3 cm. De standaardafwijking is 7,74 cm.
Tussen welke twee lengtes zit volgens de vuistregels 68% van de Nederlandse mannen? En 95%?

Antwoord

Volgens de vuistregels zit 68% van deze mannen tussen de gemiddelde lengte min de standaardafwijking en de gemiddelde lengte plus de standaardafwijking. Dus 68% heeft een lengte tussen 172,6 cm en 188,0 cm.

De 95%-regel zegt dat 95% van de lengtes maximaal twee keer de standaardafwijking van het gemiddelde af zit. Dus 95% van de mannen heeft een lengte tussen 164,8 cm en 195,8 cm.

Opgave 4

Bekijk **Voorbeeld 1**. Neem aan dat voor de verdeling van lengte L van de Nederlandse mannen boven de 20 jaar deze verdeling geldt.

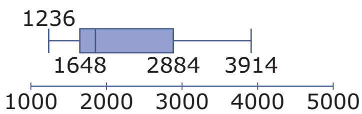
lengte Nederlandse mannen boven 20 jaar	
lengte	*percentage*
< 163	1
$163- < 168$	3,3
$168- < 173$	11,8
$173- < 178$	18,2
$178- < 183$	26,7
$183- < 188$	23,2
$188- < 193$	9,3
$193- < 198$	4,9
> 198	1,5

Tabel 5.1

Teken een bijpassend frequentiepolygoon en reken het gemiddelde $\overline{L}$ en de standaardafwijking σ_L na. Laat zien dat er sprake is van een klokvorm.

Opgave 5

Voor een onderzoek naar de levensduur van een bepaald type batterijen is op basis van 200 waarnemingen een boxplot getekend.



Figuur 5.4

Geef aan welke van de volgende beweringen waar zijn. Licht je antwoord toe.

A. Minimaal 25% van de batterijen gaat langer dan 3000 uur mee.
B. Meer dan 50% van de batterijen heeft een levensduur van minder dan 2000 uur.
C. De batterijen gaan gegarandeerd 1200 uur mee.
D. Minstens 75% van de batterijen werkt nog na 1600 uur.

Opgave 6

Op de verpakking van een pak koffie staat een inhoud van 250 gram. In werkelijkheid is dat iets meer of minder. Het gewicht van 1000 pakken koffie wordt gemeten, zonder verpakking. Uit de metingen blijkt een gemiddeld gewicht van 254 gram. De standaardafwijking is 4 gram. Ga ervan uit dat de verdeling van het gewicht klokvormig is.

Geef aan welke uitspraken volgens de vuistregels waar zijn.

A. Ongeveer 95% van de pakken koffie heeft een gewicht tussen 246 en 262 gram.
B. Ongeveer 5% van de pakken koffie heeft een gewicht onder 246 gram.
C. Ongeveer 16% van de pakken koffie heeft minder dan de beloofde 250 gram inhoud.

D. Ongeveer 50% van de pakken koffie heeft een gewicht van 250 gram.

E. Minimaal 75% van de pakken koffie heeft een gewicht van meer dan 250 gram.

Opgave 7

Een bedrijf heeft 25 werknemers in vaste dienst met een volledige werkweek. De nettoweeklonen van deze werknemers zijn verwerkt in de frequentietabel.

netto weekloon (€)	*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

Tabel 5.2

a Bepaal het gemiddelde en de standaardafwijking van deze verdeling in euro's nauwkeurig.

b Maak bij de frequentietabel een relatieve cumulatieve frequentietabel en teken de bijbehorende kleiner-gelijk-kromme (cumulatieve frequentiepolygoon).

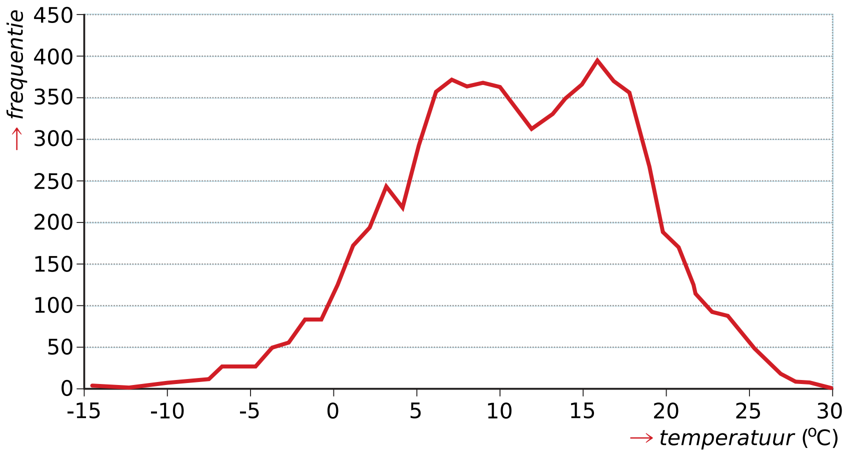
c Gebruik de relatieve cumulatieve frequentiepolygoon om het gemiddelde af te lezen. Geef aan hoe je dat doet.

d Bepaal met behulp van de kleiner-gelijk-kromme de standaardafwijking. Geef aan hoe je dat doet. Vergelijk je antwoord met dat in a.

e Hoeveel procent van alle weeklonen wijkt meer dan twee standaarddeviaties af van het gemiddelde? Klopt dit met de vuistregels voor een normale verdeling?

Voorbeeld 2

Elk uur wordt in De Bilt de temperatuur gemeten. De resultaten van acht uur ’s ochtends in de jaren 1981 tot en met 2000 zijn weergegeven in een frequentiepolygoon.



Figuur 5.5

Je kunt de vuistregels niet gebruiken voor deze frequentieverdeling, omdat het geen klokvormige verdeling is.
Geef twee afwijkingen van deze verdeling ten opzichte van een klokvormige verdeling.

Antwoord

Een klokvormige verdeling heeft één maximum, deze polygoon heeft er twee. In een klokvormige verdeling verwacht je het maximum bij de middelste waarde. Deze polygoon heeft meerdere maxima.

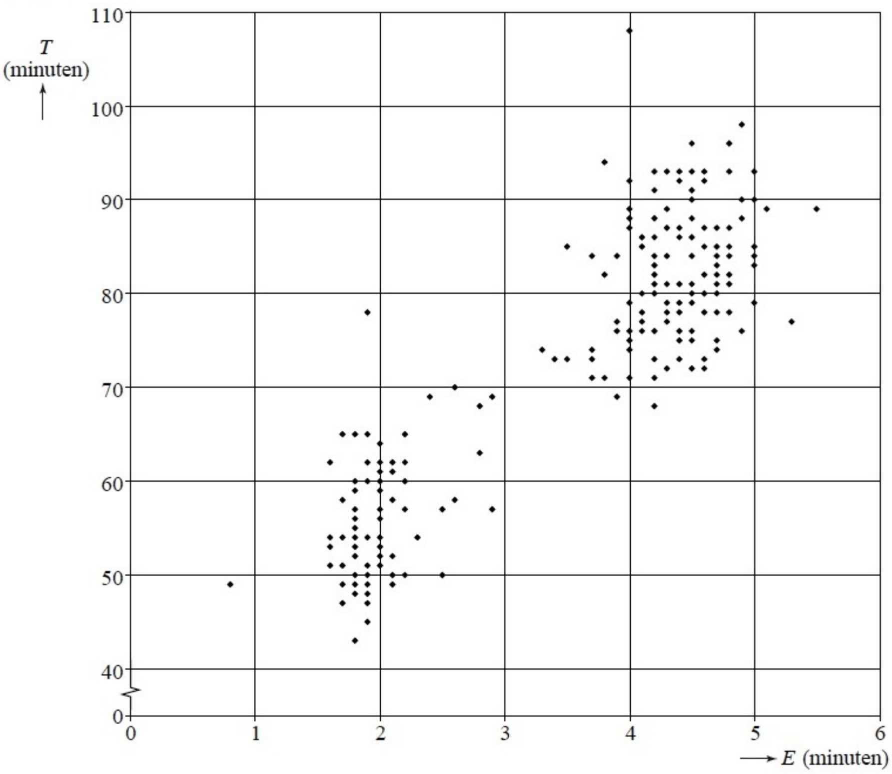
Opgave 8

Bekijk de frequentiepolygoon in **Voorbeeld 2**.

Je kunt wel de uitspraak doen dat deze frequentieverdeling scheef naar rechts verdeeld is en tweetoppig is. Leg uit.

Opgave 9

Old Faithful is de beroemdste en actiefste geiser in het Amerikaanse Yellowstone National Park (Wyoming). In een uitbarsting kan de geiser tot 32000 liter kokend water zo'n 56 m hoog spuiten. Een uitbarsting duurt tussen de $1{,}5$ en 6 minuten. De toeschouwer moet soms wel geduld hebben, want de tussenpozen tussen twee uitbarstingen variëren tussen drie kwartier en twee uur. Van 272 erupties van de Old Faithful is de duur (E) en de wachttijd vanaf de vorige eruptie (T) gegeven, beide in minuten. De verdelingen van de duur en de wachttijd zijn allebei tweetoppig. Hoe zie je dat in dit diagram?



Figuur 5.6

Voorbeeld 3

Voorbeelden van misleidingen:

Belgen spreken langzamer dan Nederlanders

De schok was groot toen uit een artikel in Onze Taal bleek dat Belgen beduidend langzamer praten dan Nederlanders. In de Randstad haalt men 5,42 lettergrepen per seconde, in Oost-Vlaanderen slechts 4,43. Sommige mensen gingen het meteen controleren. De spreeksnelheid van 21 miljoen Nederlandssprekenden werd bepaald door maar liefst 160 leraren en leraressen een stukje te laten opzeggen. Er waren acht groepen, dus twintig sprekers per groep. En dan werd ook nog gerapporteerd over het verschil tussen jong en oud, man en vrouw. Eén oude Antwerpse stotteraarster, en de achterstand is hopeloos. Onderzoeker Guy De Pauw maakte het allemaal nog erger door een dag later te verklaren dat de verschillen niet significant waren. Alsof dat er nog toe doet, met zo'n steekproef.

bron: vanmaanen.org, 2004, artikel Onze Taal, Hans van Maanen

Tabel 5.3

Vitalinea misleidt consument

In de nieuwe reclamespots voor het aanprijzen van Vitalinea van Danone gebruiken de reclame boys wel heel trieste, misleidende statistieken, waar de fouten zo van af druipen. De reclame claimt dat tijdens een studie bij 400 Belgen, is gebleken dat 80% van de deelnemers gemiddeld 3,6 kilogram afvalt. Misschien valt je frank niet direct, maar deze statistiek wil helemaal niks zeggen. Waarom geven de onderzoekers het gemiddelde gewichtsverlies van slechts 80% van de deelnemers? Waarom niet van de volle 100%? Waar zijn de statistieken van die andere 20% deelnemers plots heen? Wat mij betreft zijn deze 20% mensen die niet meegeteld zijn allemaal 30 kilogram bijgekomen door het eten van Vitalinea, en komt het gemiddelde dus uit op een gewichtstoename bij het eten van Vitalinea. Het enige dat ik kan concluderen van de reclame, is dat als je wil vermageren, Vitalinea niet het goede product is. Simpele logica.

bron: anthony.liekens.net, 2005, Anthony Liekens

Tabel 5.4

In de tekst zie je hoe slordige statistieken je kunnen misleiden en/of hoe soms slordige conclusies worden getrokken. Geef bij ‘Vitalinea misleidt consument’ kort commentaar.

Antwoord

Vitalinea misleidt consument.

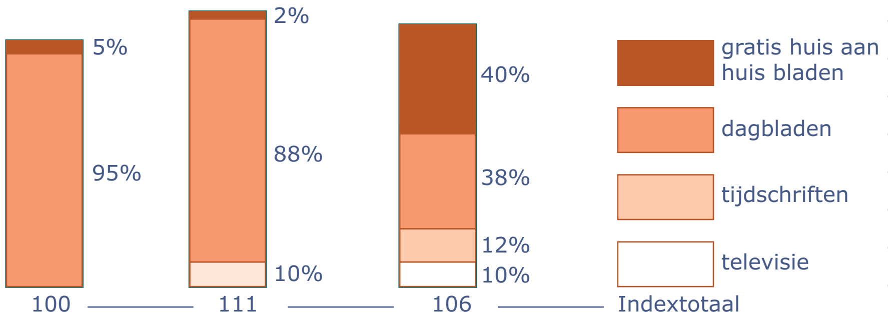
- Er wordt al aangegeven dat het vreemd is dat we niets horen over de andere 20% van de deelnemers aan het onderzoek.
- Bovendien wil het niet zeggen dat 80% van de deelnemers is afgevallen door Vitalinea. Misschien zijn ze gezonder gaan leven?

Opgave 10

In **Voorbeeld 3** zie je ook hoe in het artikel 'Belgen spreker langzamer dan Nederlanders' slordige conclusies worden getrokken. Geef kort commentaar.

Opgave 11

Je ziet waaraan een bedrijf in 2007, 2008 en 2009 reclamegelden besteedde. In 2007 werd € 150000,00 aan reclame uitgegeven.



Figuur 5.7

a Bereken hoeveel het bedrijf in 2008 aan reclame heeft uitgegeven. Bereken ook de reclameuitgaven in 2009.

b Mag je uit de gegevens concluderen dat

- het bedrijf in 2008 11% meer uitgaf aan reclame dan in 2007?
- het bedrijf in 2009 5% minder uitgaf aan reclame dan in 2008?

Licht je antwoord toe.

c Bereken hoeveel het bedrijf over de drie jaar in totaal uitgaf aan reclame in huis-aan-huisbladen. Hoeveel is dat gemiddeld per jaar?

d Vergelijk de uitgaven aan reclame in dagbladen in 2007 met die in 2008. Met hoeveel procent zijn deze uitgaven veranderd?

e Welke van deze conclusies is waar?

- In 2008 is meer aan reclame in dagbladen besteed dan in 2009 besteed is aan de totale reclame uitgezonderd de tv-reclame.
- In 2009 is meer dan € 150000,00 uitgegeven aan reclame in drukvorm.

Verwerken

Opgave 12

Bekijk het bestand **Gegevens 36 mannen**.

Op het moment van deze steekproef van 36 was de gemiddelde lengte van alle Nederlandse mannen 177,6 cm met een standaardafwijking van 6,6 cm. De verdeling van deze lengtes had een zuivere klokvorm.

gegevens van 36 mannen											
nr	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)	*nr*	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)	nr	*leeftijd* (j)	*lengte* (cm)	*gewicht* (kg)
i	L	l	G	i	L	l	G	i	L	l	G
1	20	180	95	13	46	180	95	25	67	171	68
2	23	184,5	91	14	47	177,5	75	26	67	170	81
3	23	180,5	90	15	51	196	86	27	68	169	97
4	29	169	91	16	52	188,5	93	28	68	176	72
5	30	176	80	17	53	170	85	29	71	171	84
6	33	182	82	18	54	178,5	77	30	71	180,5	75
7	36	182,5	85	19	56	172,5	74	31	71	172	77
8	36	176,5	74	20	61	164	85	32	74	171	77
9	38	180	73	21	61	181	89	33	75	183	95
10	40	190,5	112	22	61	173	95	34	75	172,5	74
11	41	184	89	23	61	181	89	35	82	173	66
12	44	179	75	24	63	181,5	100	36	84	175	90

Tabel 5.5

a Tussen welke twee waarden zou 68% van de lengtes van de 36 mannen moeten liggen als de steekproef representatief is?

b Ga na of deze steekproef voldoet aan de eerste vuistregel van de klokvormige frequentieverdeling.

c Tussen welke twee grenzen zou ongeveer 95% van de lengtes van de 36 mannen moeten liggen als de steekproef representatief is?

d Ga na of deze steekproef voldoet aan de tweede vuistregel van de klokvormige frequentieverdeling. Bekijk daarvoor de tabel met gegevens van 36 mannen.

Opgave 13

Een kledingzaak maakt broeken die op maat afgeknipt worden. Zo zijn ze voor iedereen precies lang genoeg. Maar hoe langer de broek vóór het afknippen, hoe duurder en hoe meer stof er wordt weggegooid. De ontwerpers hebben laten onderzoeken wat de beenlengte van Nederlandse mannen is. Die is klokvormig verdeeld met gemiddelde 80 cm en standaardafwijking 5 cm. De langst gemeten lengte is 95 cm.

a Alle broeken kunnen zo lang worden als de langste man nodig heeft. Hoe lang worden de broeken dan? Noem een bezwaar tegen dit idee.

Een andere mogelijkheid is om de lengte van de broek zo te kiezen dat 84% van de mannen erin past.

b Hoe lang moet de broek dan worden?

De damesafdeling van de kledingzaak wil ook zulke broeken. De verdeling van de beenlengtes van vrouwen is ook klokvormig, met $\overline{x} = 74$ cm en $\sigma_x = 4$ cm.

c Hoe lang moeten de vrouwenbroeken zijn zodat 84% van de vrouwen ze past?

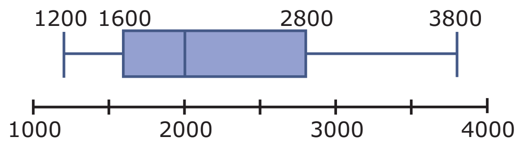
Opgave 14

In een bedrijf is het modale salaris ongeveer € 1600,00 per maand. Het gemiddelde salaris is € 1800,00 per maand. Het hoogste salaris is dat van de algemeen directeur. In de boxplot zie je de verdeling van de salarissen over alle 120 mensen die bij het bedrijf werken. Bereken in de volgende gevallen steeds het modale salaris en het gemiddelde salaris en teken de nieuwe boxplot. Doe voor elk van de drie situaties een kenmerkende uitspraak over de gevolgen van de maatregel voor de laagstbetaalde 25% werknemers.

a Alle medewerkers krijgen een loonsverhoging van 3%.

b Alle medewerkers krijgen een maandelijkse toeslag van € 200,00.

c Het salaris van de algemeen directeur wordt met € 800,00 per maand verhoogd.



Figuur 5.8

Opgave 15

Bekijk de gegevens van pasgeboren kinderen in Nederland. De verdeling is klokvormig. Doe vier uitspraken met behulp van de vuistregels over geboortegewicht en geboortelengte.

Geboortelengte in cm							
< 45,5	45,5 -< 47,5	47,5 -< 49,5	49,5 -< 51,5	51,5 -< 53,5	53,5 -< 55,5	> 55,5	
3,7	6,6	18,1	37,5	22,9	6,9	4,2	
Geboortegewicht in kg							
< 1,5	1,5 -< 2,0	2,0 -< 2,5	2,5 -< 3,0	3,0 -< 3,5	3,5 -< 4,0	4,0 -< 4,5	> 4,5
0,8	1,0	3,4	12,4	32,2	33,6	12,6	4,0

Figuur 5.9

Opgave 16

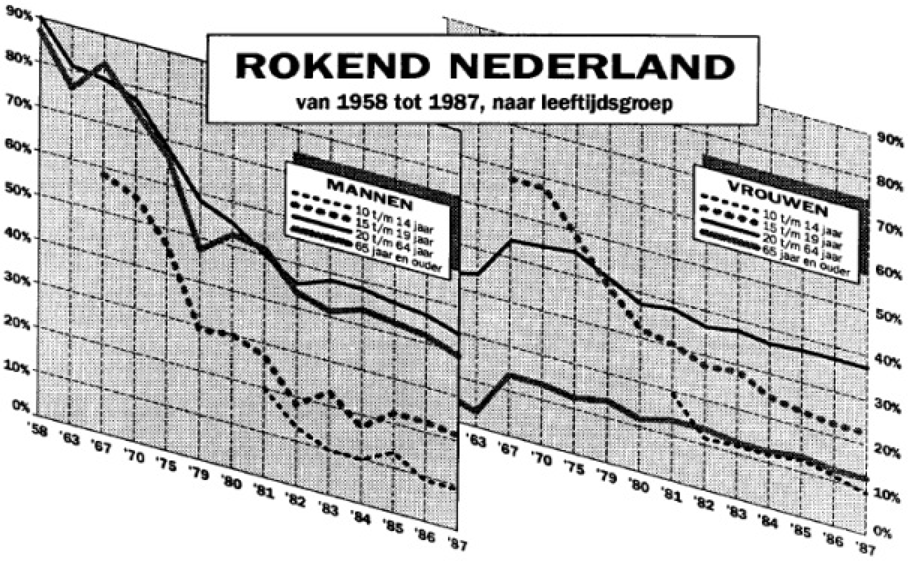
Open het Excel bestand **Patiëntgegevens**.

a Bereken de gemiddelde lengte van zowel de vrouwelijke als de mannelijke patiënten en de bijbehorende standaardafwijkingen. Is er verschil tussen de lengtes van mannen en vrouwen?

b Onderzoek of 50% van de mannen langer is dan de 84% kortste vrouwelijke patiënten.

Opgave 17

De lijndiagrammen komen uit een krantenartikel uit 1988. Volgens de linker grafiek rookte in 1958 nog 90% van de mannen in de leeftijdsgroep van 20 tot 65 jaar. In 1987 was dit percentage gedaald tot 43%. Deze sterke daling wordt door de tekenaar op een misleidende wijze benadrukt.



Figuur 5.10

a Wat veroorzaakt deze misleiding?

b Bekijk het diagram van de mannen van 15 tot 20 jaar. De grafiek ziet er ook voor de jaren 1982 tot 1987 dalend uit. Daalt het percentage rokers van die categorie ook werkelijk?

c Bij welke van deze acht diagrammen is er vrijwel nooit van daling sprake?

In het krantenartikel stond:

Een overzicht van de rookgewoonten in Nederland in 1987 gaf, net als in de jaren daarvoor, opnieuw een daling te zien van het aantal rokers in ons land. Hoewel de betrekkelijk snelle daling in de jaren zeventig en het begin van de jaren tachtig is afgenomen, heeft die tendens zich de afgelopen drie jaar gestabiliseerd op een daling van 1% per jaar. Kon in 1958 worden becijferd dat 60% van de Nederlandse mannen en vrouwen in de leeftijdsgroep van 15 tot 65 jaar rookte, volgens cijfers van de Stichting Volksgezondheid en Roken was dat in 1987 afgenomen tot 37%.

Een lezer van dit artikel denkt dat die 37% niet kan kloppen. Hij redeneert zo:

- de laatste drie jaar was er een daling van 1%;
- volgens de tekst en de figuur was de daling in de periode daarvoor nog sterker;
- in 1958 was het percentage rokers 60;
- in de 29 jaar van de periode 1958-1987 is daar zeker $29 \cdot 1\% = 29\%$ van af gegaan, dus in 1987 moet het percentage minder dan 31% zijn.

d Leg uit waarom het percentage van 37% wel correct kan zijn als je de 1% daling per jaar goed interpreteert.

Toepassen

In 1947 zijn bij 5001 vrouwelijke klanten van de Bijenkorf vijftien lichaamsmaten opgemeten. Vervolgens is gekeken welke van deze maten het meest bruikbaar zijn om een **maatsysteem voor kleding** op te baseren. Bekijk een deel van de uitkomst van het onderzoek in het bestand **Statistiek Bijenkorf 1947**.

Je kunt onder andere met behulp van de **vuistregels voor klokvormige verdelingen** uitspraken doen over deze populatie. En (er van uitgaande dat deze steekproef representatief was voor de bezoeksters van de Bijenkorf in die tijd) een maatsysteem voor vrouwenkleding bedenken.

Opgave 18

Bekijk de uitkomst van het onderzoek in het bestand **Statistiek Bijenkorf 1947**.

a Bepaal de gemiddelde lengte en de bijbehorende standaardafwijking.

b Maak het bijpassende histogram voor de variabele *lengte*.
Laat zien dat er van een mooie klokvormige frequentieverdeling sprake is.

c Hoeveel lengtes verschillen meer dan één keer de standaarddeviatie van het gemiddelde? Hoeveel procent van de vrouwen betreft dit?

d Hoeveel procent van de lengtes verschilt meer dan twee keer de standaardafwijking met het gemiddelde?

e Komen deze antwoorden overeen met de vuistregels voor klokvormige verdelingen?

f Waarom heeft het geen zin om na te gaan of voor de frequentieverdeling van de gewichten de vuistregels opgaan?

Testen

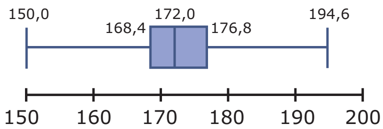
Opgave 19

Je ziet een boxplot van de lengtes van 1064 vaders van ongeveer 100 jaar geleden.

a Welke uitspraak kun je doen over de 25% kortste mannen?

b Controleer de uitspraak: 266 van deze 1064 mannen had een lengte vanaf 172,0 tot 176,8 cm.

c Geef twee redenen waarom de lengtes van deze mannen geen zuiver klokvormige frequentieverdeling hebben.



Figuur 5.11

Opgave 20

Je ziet de leeftijdsopbouw van leraren in het havo/vwo in procenten.

a Bereken voor elk van de vijf genoemde jaren het gemiddelde en de standaarddeviatie van de leeftijden van deze leraren.

b Teken de vijf frequentiepolygonen en geef die waarden daarin aan.

c Welke conclusies kun je trekken?

d De waarden van 1995 en 2000 zijn schattingen die de onderzoekers in 1994 hebben gedaan. Passen die schattingen bij de gegevens uit de voorgaande jaren?

Leraren in HAVO/VWO					
Leeftijd	*1980*	*1985*	*1990*	*1995*	*2000*
-29	20	12	8	3	6
30-34	21	19	13	8	4
35-39	18	21	19	14	9
40-44	14	18	21	20	14
45-49	11	13	19	22	20
50-54	8	10	14	18	22
55-59	5	6	5	12	17
60-64	3	1	1	3	8

Tabel 5.6

(bron: 'Onderwijswacht Gelders onderzoek' - Arbon 1994)

Opgave 21

Open het bestand **Etmaaltemperaturen De Bilt**.

a Maak een histogram van de temperaturen in de maand juli over de jaren 1755 tot 1900. Neem een klassenbreedte van 1 °C.

b Maak ook een histogram voor de periode van 1900 tot 2007.

c Vergelijk de twee histogrammen met elkaar. Kun je concluderen dat de temperatuur in de maand juli na 1900 gemiddeld hoger is dan in de voorgaande periode?

2.6 Totaalbeeld

Samenvatten

Je hebt nu het onderwerp **Statistiek** doorgewerkt. Er moet een totaalbeeld van deze leerstof ontstaan...
Ga na, of je al de bij dit onderwerp horende begrippen kent en weet wat je er mee kunt doen. Ga ook na of je de activiteiten die staan genoemd kunt uitvoeren. Maak een eigen samenvatting!

Begrippenlijst

- kwalitatieve variabelen — discrete/continue kwantitatieve variabelen
- (relatieve) frequentie — frequentietabel, frequentieverdeling — somfrequenties — klassenindeling
- staafdiagram, lijndiagram, cirkeldiagram — histogram, (som-) frequentiepolygoon — steelbladdiagram
- centrummaten: modus, mediaan, gemiddelde — spreidingsmaten: spreidingsbreedte, kwartielafstand, standaardafwijking — kwartielen, boxplot
- klokvormige frequentieverdeling — twee vuistregels

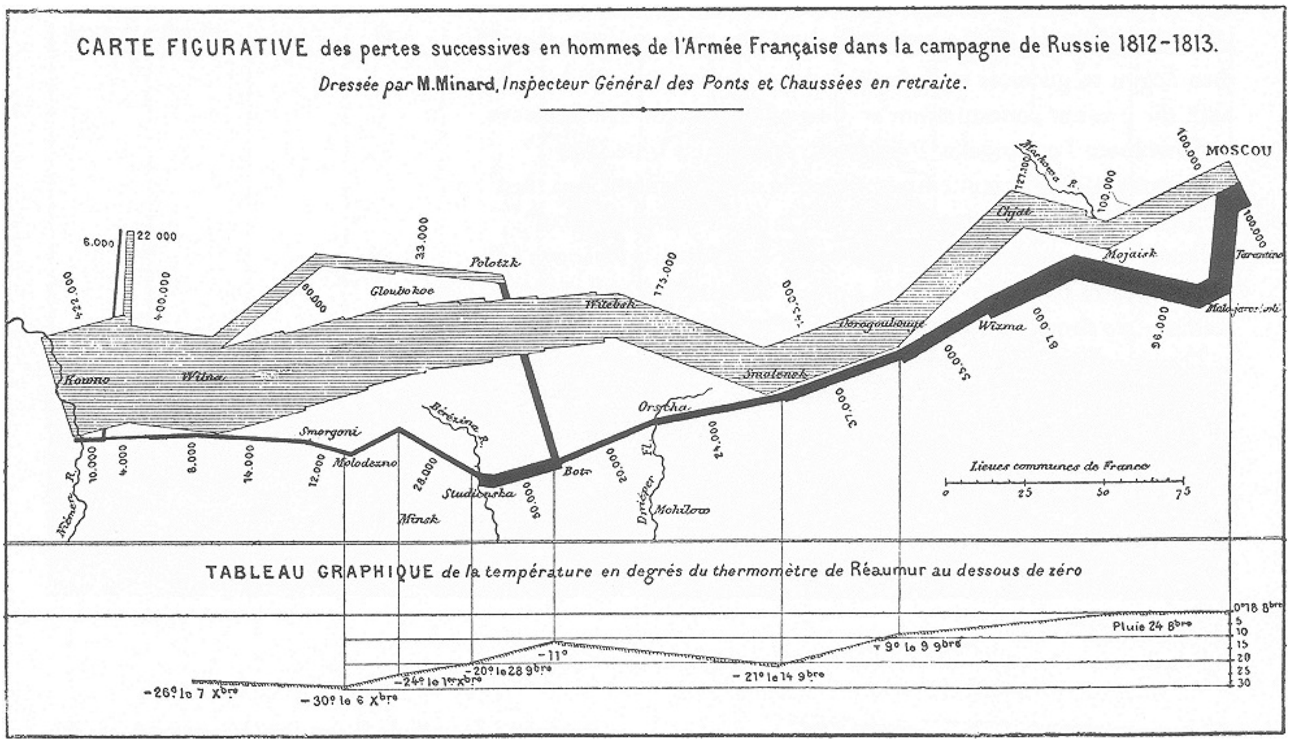
Activiteitenlijst

- representativiteit van een steekproef beoordelen — soorten statistische variabelen herkennen
- (som)frequentieverdelingen maken — werken met klassen
- diagrammen gebruiken om statistische resultaten weer te geven (ook met de GR en/of Excel)
- centrummaten en spreidingsmaten berekenen (schatten) zowel met als zonder klassenindeling (ook met de GR en/of Excel)
- uitspraken doen op grond van steekproefverdelingen

Achtergronden

De beschrijvende statistiek ontstond pas toen staten en centrale overheden een belangrijke rol gingen spelen. Immers: vooral centrale overheden zijn geïnteresseerd in globale overzichten van de bevolking, omdat het aantal mensen waarover zij verantwoordelijk is te groot is om van iedereen alles te weten.
Via **Historische hoogtepunten van de grafische verwerking** kun je veel informatie vinden over het ontstaan van statistieken en de mensen die daarin een rol hebben gespeeld.

Hieronder zie je Minard's beroemde diagram van de tocht van Napoleon naar Rusland.



Figuur 6.1

Testen

Opgave 1

In 1987 verscheen het geruchtmakende boek ‘Women and Love: A Cultural Revolution in Progress’ van Shere Hite. De auteur beschreef daarin de resultaten van een onderzoek onder 100000 vrouwen met betrekking tot hun relatie. Van de vrouwen die de vragenlijst terugstuurden:

- voelde 84% zich emotioneel niet goed in hun relatie;
- had 95% psychische of lichamelijke mishandeling doorstaan;
- gaf slechts 13% aan na twee jaar huwelijk nog van hun man te houden.

Om te laten zien dat Hite een representatieve steekproef had opgesteld, gebruikte ze tabellen. Hite hield onder meer rekening met ras/afkomst en de regio. Zo’n 4500 vrouwen stuurden de vragenlijst ingevuld terug naar de auteur.

a Specificeer zo nauwkeurig mogelijk welke populatie Shere Hite wilde onderzoeken met behulp van deze steekproef en geef aan met welk type statistiek Shere Hite bezig was. Met beschrijvende of met verklarende statistiek?

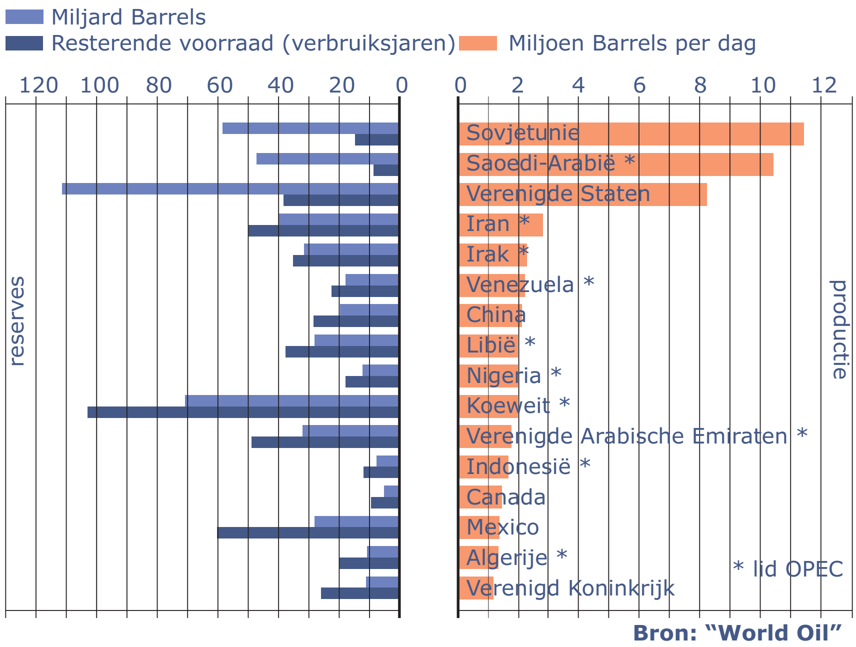
b Shere Hite heeft rekening gehouden met jaarlijks inkomen, type woonplaats, regio en met ras/afkomst van de onderzochte vrouwen. Op welke twee kenmerken had ze haar steekproef nog verder kunnen uitsplitsen om de steekproef nog representatiever te maken?

c Uit de gegevens die je nu hebt over dit onderzoek kun je niet afleiden of deze steekproef aselect is. Definieer de term aselecte steekproef voor dit onderzoek.

d Kon Shere Hite met 100% zekerheid weten dat 87% van de vrouwen na twee jaar huwelijk niet meer van hun man hield? Licht je antwoord toe.

Opgave 2

In de figuur vind je gegevens over de oliereserves en de olieproductie uit ‘Aardolie, de halve wereld draait er op’, van de Stichting School en Bedrijf.



Figuur 6.2

a Voor welk type variabele zie je staafdiagrammen omtrent olie afgebeeld?

b Hoeveel procent van de reserves, gerekend in vaten, is in handen van de OPEC-landen?

c Wat is de productieproportie van Rusland (vroeger: Sovjetunie)? Geef je antwoord in procenten nauwkeurig.

d Geef zowel de reserves (in miljarden vaten) als de productie (in miljoenen vaten per dag) weer in een cirkeldiagram. Neem als categorieën de OPEC-landen, de Verenigde Staten, Rusland en overig.

e Vergelijk de oliereserves en de olieproductie met elkaar op basis van je cirkeldiagrammen.

Opgave 3

Op de verpakking van een literpak melk staat ‘Inhoud 1 liter’. In werkelijkheid wil dat nog wel eens iets meer of minder zijn. Uit metingen blijkt een gemiddelde inhoud van 1,002 liter. De standaardafwijking is 0,004 liter. De verdeling van de inhoud is klokvormig. Geef met behulp van de vuistregels bij de uitspraken aan of ze waar of niet waar zijn.

a Ongeveer 95% van de pakken melk heeft een inhoud van meer dan 1,010 liter.

b Ongeveer 32% van de pakken melk bevat minder dan 0,998 of meer dan 1,006 liter.

c Ongeveer 50% van de pakken melk bevat $1,002$ liter.

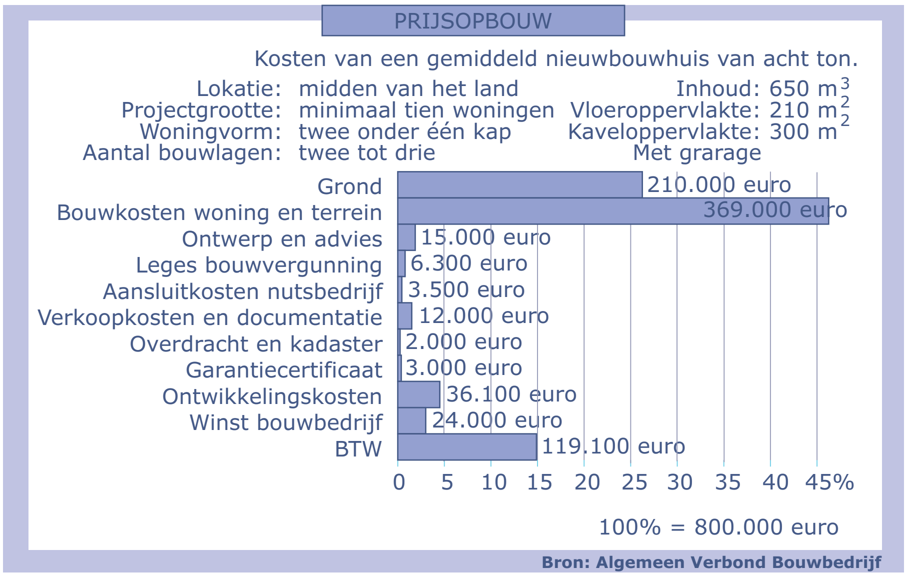
Stel je voor dat er iets mis is met de vulmachine van de melkpakken en dat de frequentieverdeling van de inhoud opeens scheef naar rechts wordt in plaats van klokvormig.

d Beargumenteer of de volgende uitspraak wel of niet met zekerheid kan worden gedaan:

'Méér dan 50% van de melkpakken heeft een inhoud die groter is dan het (nieuwe) gemiddelde.'

Opgave 4

In de figuur wordt aangegeven hoe de prijs van een nieuwbouwhuis is opgebouwd.



Figuur 6.3

a Hoeveel is de grondprijs per vierkante meter bij deze woningen?

b Bereken hoeveel procent van de totale kosten opgaan aan bouwkosten woning en terrein.

c Maak een cirkeldiagram van de prijsopbouw van deze woningen. Werk met vier categorieën, te weten: grond, bouwkosten, btw en overige.

d Het hoge btw-tarief was toen 19%. In de figuur zie je dat aan btw ongeveer 15% moet worden betaald. Geef een verklaring voor dit verschil.

(bron: Bouwend Nederland, 2000)

Opgave 5

De tabel geeft de leeftijden van het personeel van twee bedrijven.

leeftijden van het personeel			
bedrijf A		bedrijf B	
leeftijd	*aantal*	*leeftijd*	*aantal*
$15- < 18$	89	$15- < 18$	67
$18- < 25$	133	$18- < 25$	205
$25- < 40$	414	$25- < 40$	801
$40- < 55$	249	$40- < 55$	562
$55- < 60$	68	$55- < 60$	265
$60- < 65$	47	$60- < 65$	100
	1000		2000

Tabel 6.1

a De twee bedrijven willen fuseren. Werknemers vanaf 60 jaar kunnen gebruikmaken van een afvloeiingsregeling.
Hoeveel procent van de werknemers van het gefuseerde bedrijf kan daar gebruik van maken?

b Wat is er (theoretisch) mis met de frequentietabellen van bedrijf A en bedrijf B?

c Leg uit waarom een cumulatieve relatieve frequentiepolygoon van dergelijke frequentietabellen de zuiverste indruk geeft. Waarom geven een (relatieve) frequentiepolygoon, een histogram en een cirkeldiagram hier een vertekend beeld?

d Wil je deze twee frequentietabellen alleen gebruiken om ze met elkaar te vergelijken, dan doen de bezwaren er minder toe. Teken de relatieve histogrammen van beide bedrijven met gebruikmaking van de gegeven klassenindeling en benoem de verschillen en overeenkomsten tussen beide frequentieverdelingen die je uit deze histogrammen mag en kunt afleiden.

Opgave 6

Eind vorige eeuw bleek uit onderzoek van het gemengde boerenbedrijf het houden van kippen een belangrijke rol te spelen bij het tot stand komen van het inkomen van deze boeren. Daarom werd de boeren gevraagd naar het aantal kippen op hun bedrijf.

aantal kippen	*aantal bedrijven*	*aantal kippen*	*aantal bedrijven*
$1-10$	5	$101-110$	123
$11-20$	12	$111-120$	101
$21-30$	19	$121-130$	85
$31-40$	24	$131-140$	79
$41-50$	33	$141-150$	60
$51-60$	52	$151-160$	43
$61-70$	69	$161-170$	21
$71-80$	75	$171-180$	9
$81-90$	108	$181-190$	4
$91-100$	120	$191-200$	2

Tabel 6.2

a Met welk type variabele heb je hier te maken?

b Teken een cumulatief relatief frequentiepolygoon.

c Schat de mediaan en de beide kwartielen. Teken een boxplot bij deze gegevens.

d Je kunt het gemiddelde en de standaarddeviatie schatten vanuit de klassenmiddens en de frequentietabel. Laat zien hoe dat gaat en geef je antwoord in één decimaal nauwkeurig.

e Als het aantal kippen op een gemengd boerenbedrijf een klokvormige verdeling kent, hoeveel kippen hebben dan de 2,5% gemengde boerenbedrijven met de meeste kippen?

Toepassen

Opgave 7: Coopertest

In de sport wordt veel met statistieken gewerkt. Er wordt namelijk nogal wat gemeten...
De Cooper-test is een bekende manier om de conditie te meten. Je meet dan hoever je kunt hardlopen in 12 minuten. En de afgelegde afstand zegt iets over je fitheid.

Dit **XL-bestand met Coopertest resultaten** van een groep scholieren levert een verzameling basisgegevens. In deze tabel zie je hoeveel iemand onder de 30 jaar kan lopen met een conditie die slecht / redelijk / goed / zeer goed is.

	vrouwen	mannen
slecht	$0- < 1800$	$0- < 2000$
redelijk	$1800- < 2200$	$2000- < 2400$
goed	$2200- < 2700$	$2400- < 2800$
zeer goed	> 2700	> 2800

Tabel 6.3

Nu kun je nagaan hoeveel procent van deze groep scholieren in welke categorie valt, voor mannen en vrouwen afzonderlijk of voor beide groepen samen. Ook kun je de gemiddelde score, de standaardafwijking, e.d., berekenen. En nagaan of er van een klokvormige verdeling sprake is, de vuistregels kloppen...
Maar het is natuurlijk leuker om dit met resultaten van de eigen school te doen.

In 2004 volbrachten 182 leerlingen in klas 4 en 5 een Coopertest. Van de leerlingen die de Coopertest (12 minutenloop) uitliepen is de afgelegde afstand genoteerd. Misschien kun je ook werken met gegevens van je eigen school!

a Welke klassenindeling zou je kiezen voor de vrouwen en welke voor de mannen om te kunnen concluderen of de conditie slecht, redelijk, goed of zeer goed is?

b Verwerk de gegevens voor vrouwen en mannen apart in twee afzonderlijke histogrammen.

c Verwerk de gegevens voor de vrouwen in een cirkeldiagram. Verwerk de gegevens voor de mannen in een cirkeldiagram.

d Welk percentage vrouwen heeft een goede of zeer goede conditie? Welk percentage mannen heeft een goede of zeer goede conditie?

e Typeer de afgelegde afstanden van de mannen en de vrouwen bij de Coopertest in gemiddelde, modus en mediaan.

f Teken van de resultaten van de mannen en van de vrouwen een boxplot. Bereken de kwartielafstand en de spreidingsbreedte.

g Doe een uitspraak op grond van beide boxplots waarin je de conditie van de mannen en de vrouwen met elkaar vergelijkt.

h Bereken voor beide groepen de standaardafwijking.

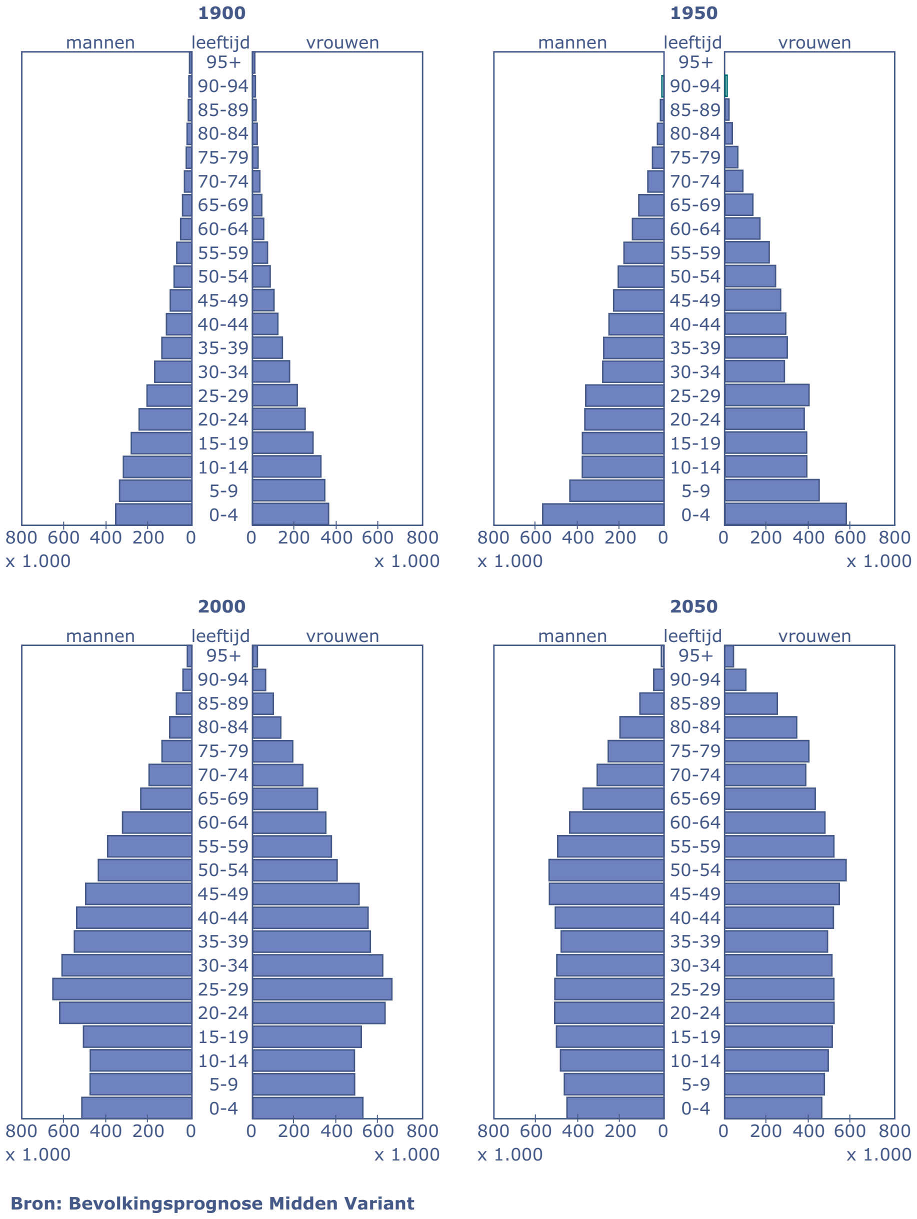
i Van hoeveel procent van de mannen en van hoeveel procent van de vrouwen wijkt de gelopen afstand meer dan de standaardafwijking van het gemiddelde af?

j Van hoeveel procent van de mannen en van hoeveel procent van de vrouwen wijkt de gelopen afstand meer dan twee keer de standaardafwijking van het gemiddelde af?

k Doe een uitspraak op grond van deze laatste gegevens waarin je de conditie van de mannen en de vrouwen met elkaar vergelijkt.

Opgave 8: Leeftijdsdiagrammen NL

Je ziet hier vier leeftijdsdiagrammen van Nederland.



Figuur 6.4

a Wat voor soort diagrammen zijn dit?

b Hoeveel kinderen van 0−4 waren er in 1900 ongeveer? En in 1950?

c In de jaren rond 1950 zijn er nogal veel kinderen geboren. Waaraan kun je dat zien?

d Waaraan kun je zien dat Nederland aan het vergrijzen is?

e Waaraan kun je zien dat vrouwen gemiddeld ouder worden dan mannen?

f De belangrijkste centrummaat bij een leeftijddiagram is de modale klasse. Welke klasse was in 1900 de modale klasse? En in 2050?

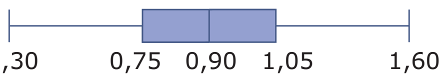
g Welke gevolgen heeft het feit dat de modale klasse steeds hoger komt te liggen?

h De bevolkingsopbouw van Nederland is ook goed zichtbaar te maken in boxplots. Teken bij het leeftijdsdiagram van 1900 een boxplot (mannen en vrouwen samen). Doe dat ook bij het geschatte leeftijdsdiagram van 2050. Noem minstens drie karakteristieke verschillen en geef er verklaringen bij.

Examen

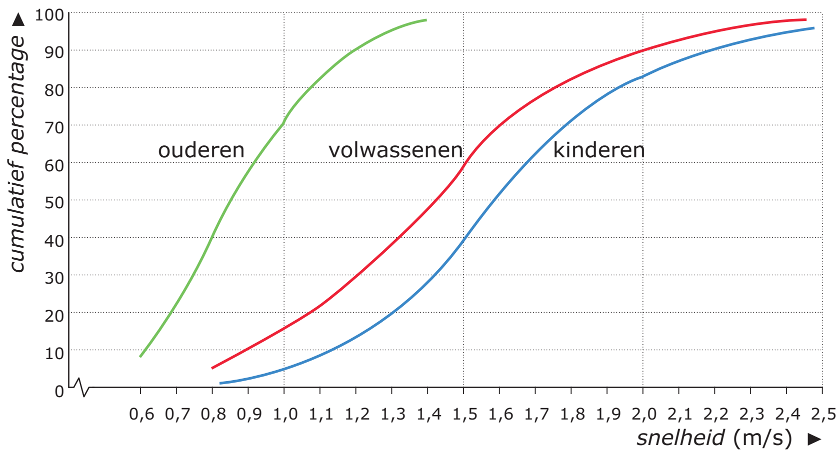
Opgave 9: Loopsnelheden van voetgangers

Men heeft onderzoek gedaan naar de loopsnelheden van voetgangers. Bij dit onderzoek zijn de voetgangers in drie leeftijdsgroepen verdeeld, namelijk kinderen, volwassenen en ouderen. Met de gegevens uit het onderzoek heeft men een boxplot gemaakt voor de loopsnelheden van de groep ouderen.



Figuur 6.5

De snelheden die bij de boxplot vermeld zijn, zijn in meters per seconde. Meer gedetailleerde informatie over de groepen zie je in de volgende figuur.



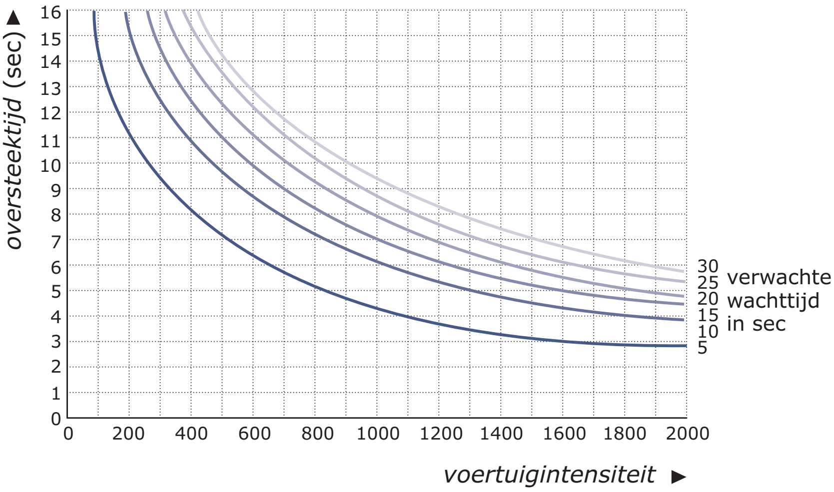
Figuur 6.6

Op de verticale as staat een cumulatief percentage; dit houdt in dat afgelezen kan worden hoeveel procent van de mensen van de verschillende groepen met de aangegeven snelheid of een lagere snelheid loopt. Zo kun je bijvoorbeeld aflezen dat voor de groep ouderen bij een snelheid van 1 m/s het cumulatieve percentage bijna 70 is. Dus bijna 70% van de ouderen loopt met een snelheid

van 1 m/s of langzamer. Aan de hand van onder andere deze gegevens wordt een model gemaakt voor de tijd die de mensen nodig hebben om een weg over te steken.
Neem aan dat de loopsnelheden ook voor het oversteken van een weg gelden. We bekijken het oversteken van een 20 meter brede weg. Er wordt recht overgestoken, dus men loopt daarbij 20 m.

a Maak met behulp van de gegevens een boxplot voor de oversteektijden van ouderen. Licht je werkwijze toe.

Tot nu toe hebben we alleen gekeken naar de tijd van oversteken zelf. Als je bij een weg aankomt, kun je niet altijd meteen oversteken; soms moet je een aantal seconden wachten. Deze wachttijd hangt samen met de drukte op de weg en de benodigde oversteektijd. De drukte op de weg wordt aangegeven met het aantal voertuigen dat per uur passeert (voertuigenintensiteit). Omdat ouderen in het algemeen minder snel lopen, zal voor deze groep de benodigde oversteektijd en dus ook de wachttijd groter zijn dan bijvoorbeeld voor kinderen. Er is een model gemaakt voor de samenhang tussen oversteektijd, voertuigenintensiteit en verwachte wachttijd.



Figuur 6.7

In de figuur hierboven is dat voor zes verschillende wachttijden in beeld gebracht. Uit deze figuur is bijvoorbeeld af te lezen dat volgens dit model bij een oversteektijd van 9 s en een voertuigenintensiteit van 700 voertuigen per uur rekening gehouden moet worden met een wachttijd van 15 s.

b Teken de grafiek die het verband aangeeft tussen de oversteektijd en de verwachte wachttijd bij een voertuigenintensiteit van 800. Teken de grafiek alleen voor wachttijden van 5 tot en met 30 s.

We willen een beeld krijgen van de totale tijd die een rol speelt bij het oversteken van een weg van 20 m breed en een voertuigenintensiteit van 800 voertuigen per uur. We spreken dan over de somtijd. Als we iemands verwachte wachttijd en zijn oversteektijd optellen, krijgen we zijn somtijd.

We bekijken nu de groep van volwassenen. De hoogste snelheid die in deze groep is waargenomen is $2{,}6$ m/s.

c Wat is de langste somtijd en wat is de kortste somtijd van de 10% snelste volwassenen? Licht je antwoord toe.

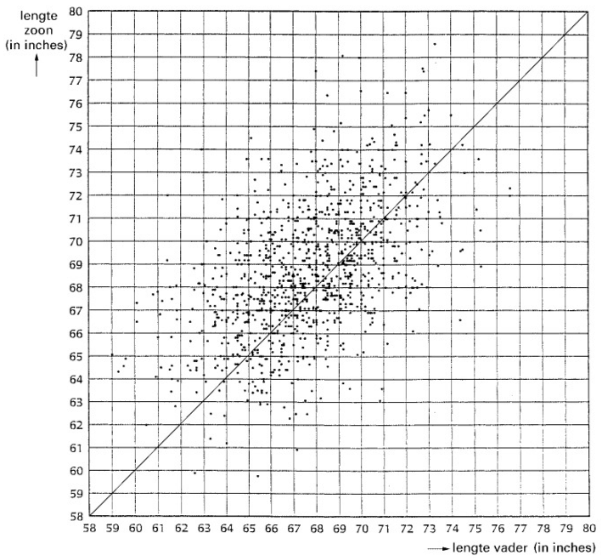
(bron: examen wiskunde A havo 1994, tweede tijdvak)

Opgave 10: Vaders en zonen

De Engelsman Karl Pearson was een van de grondleggers van de moderne statistiek. Hij heeft zich vaak bezig gehouden met statistiek over biologische onderwerpen. Ongeveer een eeuw geleden onderzocht hij, samen met zijn collega Alice Lee, of in Engeland zonen gemiddeld langer zijn dan hun vaders. Zij vergeleken de lengtes van 1064 zonen en hun vaders. De zonen studeerden allen aan een Londense universiteit.

a Is hier sprake van een aselecte steekproef? Licht je antwoord toe.

De resultaten zijn in deze figuur weergegeven. Elke stip stelt één vader-zoon-paar voor. De lengte van de vader staat op de horizontale as, de lengte van de zoon op de verticale as. De lengtes zijn gegeven in inches (1 inch = $2{,}54$ cm).



Figuur 6.8

In de figuur is een lijn getekend. Als een stip op deze lijn ligt, dan zijn de vader en de zoon precies even lang. We noemen een vader en zijn zoon ongeveer even lang als ze minder dan 2 inch in lengte verschillen.

b Teken in de figuur op **de bijlage** het gebied waarin de punten liggen die horen bij vaders en zonen die ongeveer even lang zijn. Licht je werkwijze toe.

c Kun je met behulp van het getekende gebied concluderen dat de zonen gemiddeld langer zijn dan hun vaders? Licht je antwoord toe.

Op **de bijlage** zie je een boxplot van de lengtes van de 1064 vaders. De vijf kenmerkende getallen van de boxplot staan erbij.
In **dit document** vind je ook een lijst met de lengtes van alle 1064 zonen. De getallen in deze lijst staan op volgorde van groot-

te. Na iedere 10 getallen staat een streepje. Na iedere 50 getallen staat bij het streepje hoeveel getallen er tot daar staan.

d Teken de boxplot van de lengtes van de zonen. Schrijf de vijf kenmerkende getallen van de boxplot erbij.

Het onderzoek dat bovenstaande getallen opleverde, is ongeveer honderd jaar geleden gedaan. In die tijd hadden jonge mannen een gemiddelde lengte van 68,6 inch. Dat is niet zo groot want 68,6 inch is maar 174 cm (1 inch = 2,54 cm).
Tegenwoordig is de helft van de jonge mannen langer dan 182 cm. Honderd jaar geleden was veel minder dan de helft van de jonge mannen zo lang. De lengte was toen klokvormig verdeeld met een gemiddelde van 68,6 inch en een standaardafwijking van 2,7 inch.

e Onderzoek met de vuistregels of het aantal jonge mannen langer dan 182 cm in die tijd groter of kleiner dan 16% van het totaal aantal jonge mannen was.

(bron: examen wiskunde A havo 2003, eerste tijdvak, aangepast)

Register

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

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

Stichting Math4All

Inhoud Katern 2

www.math4all.nl

Werkblad bij Opgave 10 op pagina 126.