4.1 Kansverdelingen

Inleiding

Een bedrijf is geïnteresseerd in de tijd die nodig is om een klant te helpen, de transactietijd. Je kunt die tijd in klassen indelen (bijvoorbeeld van hele minuten) en zo bijhouden hoeveel minuten een transactie duurt. Je maakt dan een kansverdeling voor de variabele transactietijd t. Maar de transactietijd kan in feite elke (positieve) reële waarde aannemen. Door steeds kleinere tijdsintervallen te nemen kun je een kansverdeling voor deze continue kansvariabele opstellen, benaderen.

Figuur 1

Je leert in dit onderwerp

- het begrip kansvariabele en bijbehorende synoniemen toevalsvariabele en statistische variabele;
- het begrip kansverdeling bij een continue variabele, in het bijzonder de normale verdeling;
- werken met (de vuistregels van) een normale kansverdeling.

Voorkennis

- de vuistregels voor klokvormige frequentieverdelingen
- centrummaten (met name gemiddelde en mediaan) en spreidingsmaten (met name de standaardafwijking)

Verkennen

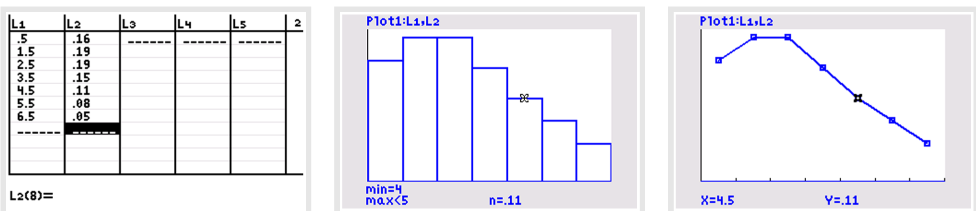
Opgave V1

Een bedrijf heeft door tellingen een frequentieverdeling opgesteld voor de tijd die nodig is om een klant te helpen. Voor deze transactietijd (in minuten) geldt:

t (min)	1	2	3	4	5	6	7	8	9	10
$P(T = t)$	0,16	0,19	0,19	0,15	0,11	0,08	0,05	0,04	0,03	0,01

Tabel 1

Je kunt dit opvatten als een kansverdeling voor de variabele T. T kan in feite elke (positieve) reële waarde aannemen. Door steeds kleinere tijdsintervallen te nemen kun je een kansverdeling voor deze continue variabele benaderen. Het lijndiagram laat dit al een beetje zien. (Merk op dat voor de juiste figuren de klassenmiddens van elke minuut moeten worden ingevoerd in de grafische rekenmachine.)



Figuur 2

a Hoe groot is de kans dat een klant hoogstens 4 minuten transactietijd kost?

b Hoe geef je die kans in het staafdiagram weer? En in het lijndiagram?

c Hoe bepaal je de kans dat een klant minder dan 4,75 minuten transactietijd kost?

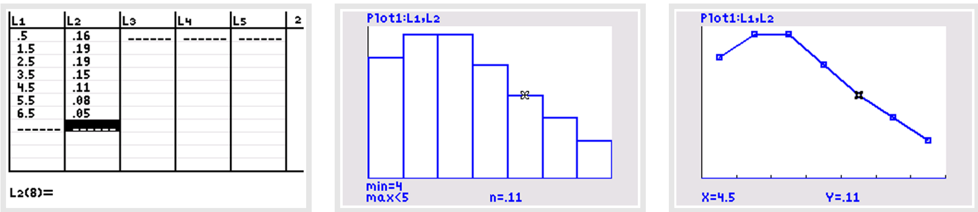
Uitleg 1

Een bedrijf noemt de tijd die nodig is om een klant te helpen de transactietijd. Het bedrijf heeft door tellen de relatieve frequenties van die transactietijd vastgesteld.

t (min)	1	2	3	4	5	6	7	8	9	10
$\text{P}(T = t)$	0,16	0,19	0,19	0,15	0,11	0,08	0,05	0,04	0,03	0,01

Tabel 2

Vat dit op als een kansverdeling voor de statistische variabele T.
De kansverdeling kun je invoeren op de grafische rekenmachine om er vervolgens een histogram of een lijndiagram bij te maken.



Figuur 3

- De kans dat een klant hoogstens vier minuten transactietijd kost, is dan:
  $\text{P}(T \leq 4) = 0{,}16 + 0{,}19 + 0{,}19 + 0{,}15 = 0{,}69$
  Dit is de oppervlakte van de eerste vier staafjes van het histogram.
  Het is ook de oppervlakte onder het lijndiagram vanaf $t = 0$ tot $t = 4$.
- De kans dat een klant hoogstens 4,75 minuten transactietijd kost, kun je benaderen door de oppervlakte te schatten onder het lijndiagram die de middens van de bovenkanten van de staafjes verbindt vanaf $t = 0$ tot $t = 4{,}75$.

T kan in werkelijkheid elke positieve waarde aannemen. Door meer metingen te doen, kun je een kansverdeling voor deze variabele opstellen met heel veel staafjes, zelfs zo veel staafjes dat de middens van hun bovenkanten een vloeiende kromme vormen.

Opgave 1

Bekijk in **Uitleg 1** de tabel met de relatieve frequentie van de transactietijd.

a Hoe groot is de kans op een transactietijd van twee minuten?

b Hoe groot is de kans op een transactietijd van meer dan acht minuten?

Opgave 2

Bekijk de diagrammen van de relatieve frequentieverdeling van de transactietijd in **Uitleg 1**.

a Waarom kun je T hier opvatten als discreet?

b Waarom is T in werkelijkheid een continue variabele?

c Hoe moet het bijbehorende kanshistogram er ongeveer uitzien als je T continu neemt?

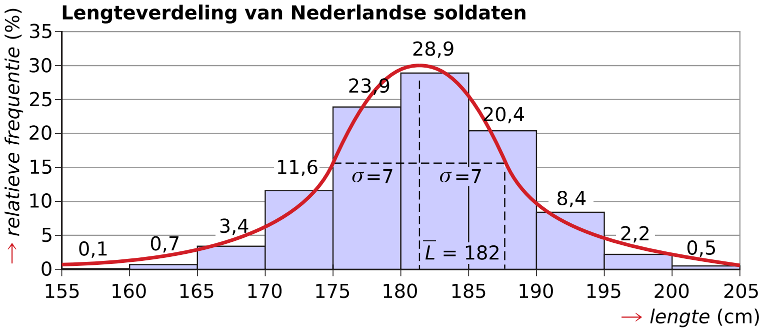
d Als je T als discrete variabele opvat, hoeveel seconden heb je dan maximaal gewacht als je in de klasse $T = 3$ minuten valt?

e Bepaal de kans $\text{P}(T \leq 3)$ als je T als discrete variabele opvat.

f Schat de kans $\text{P}(T \leq 3)$ in het geval je T opvat als continue variabele.

Uitleg 2

Bekijk het histogram van de verdeling van de lichaamslengte van een groep soldaten.



Figuur 4

De lichaamslengte L lijkt discreet door de indeling in klassen.
In werkelijkheid is de lichaamslengte een continue toevalsvariabele. In de figuur is een passende kromme getekend die de continue verdeling van de lichaamslengte L benadert. De benadering wordt beter als je meer klassen maakt.

De grafiek heeft een klokvormige frequentieverdeling die wordt bepaald door gemiddelde $\overline{L} = 182$ en standaardafwijking (ook wel standaarddeviatie genoemd) $\sigma(L) = 7$. L is normaal verdeeld. Het gemiddelde van een normale verdeling wordt μ (spreek uit 'mu') genoemd. Er geldt:

- het hoogste punt van elke normale verdeling zit bij μ.
- een maat voor de spreiding is de standaardafwijking σ.
- elke normale verdeling is symmetrisch ten opzichte van μ in het midden.
- hoe verder je bij μ vandaan gaat (naar links of naar rechts), hoe dichter de hoogte van de normale verdeling bij 0 komt.

Kansen bepaal je door de bijpassende oppervlakte onder de grafiek te schatten. De totale oppervlakte onder deze kromme is altijd 1 of 100%.
Je kunt ook de al bekende vuistregels voor de normale verdeling gebruiken om kansen te benaderen.
Van de oppervlakte onder de normaalkromme:

- ligt ongeveer 68% tussen $X = \mu - \sigma$ en $X = \mu + \sigma$
- ligt ongeveer 95% tussen $X = \mu - 2\sigma$ en $X = \mu + 2\sigma$
- ligt ongeveer 100% tussen $X = \mu - 3\sigma$ en $X = \mu + 3\sigma$

Opgave 3

Bekijk de relatieve frequentieverdeling van de lengtes van een groep soldaten op een kazerne in **Uitleg 2**.

a Waarom is L een continue toevalsvariabele?

b Welk percentage hoort bij het hele gebied onder de kromme?

c Een klokvormige verdeling is symmetrisch. Waarom volgt uit die symmetrie dat $P(L \leq 182) = 0{,}5$?

d Hoeveel is $P(175 \leq L \leq 189)$.

e Hoeveel is $P(L \leq 189)$.

f Maak een schatting van $P(L \leq 174)$.

Theorie en voorbeelden

Om te onthouden

Bij een statistische variabele waarvan de frequenties afhangen van het toeval hoort een frequentieverdeling. Bij relatieve frequenties spreek je van een **kansverdeling**.

Bij een discrete variabele kun je die weergeven door een histogram.

Bij een continue variabele kun je die weergeven door een vloeiende kromme door (de middens van) de bovenkanten van zeer veel denkbeeldige dunne staafjes.

Bij een continue variabele is de kans gelijk aan de oppervlakte op een interval onder die vloeiende kromme. De totale oppervlakte onder deze kromme is altijd 1 of 100%.

Bekijk de applet.

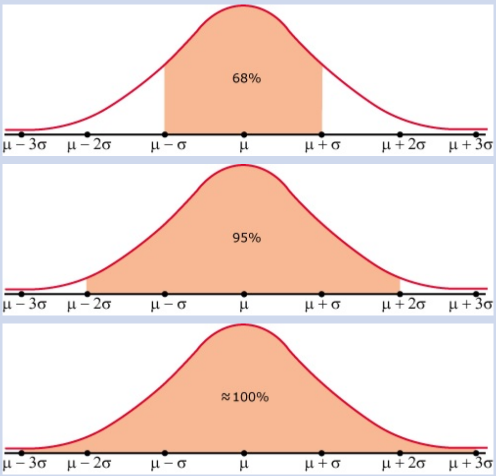
Een veelvoorkomende kansverdeling is de **normale verdeling**, de verdeling in de vorm van een klok.

Van deze kansverdeling wordt de vorm, de **normaalkromme**, volledig bepaald door het gemiddelde $\mu = \overline{X}$ en de standaardafwijking (of standaarddeviatie) σ.

Voor een normaal verdeelde variabele X geldt:

- het hoogste punt zit bij $X = \mu$.
- de spreiding is de standaardafwijking σ.
- er is symmetrie ten opzichte van de lijn $X = \mu$.
- de hoogte komt steeds dichter bij 0 als X verder van μ af ligt.
- van de oppervlakte onder de normaalkromme ligt:
  ongeveer 68% tussen $X = \mu - \sigma$ en $X = \mu + \sigma$
  ongeveer 95% tussen $X = \mu - 2\sigma$ en $X = \mu + 2\sigma$
  ongeveer 100% tussen $X = \mu - 3\sigma$ en $X = \mu + 3\sigma$

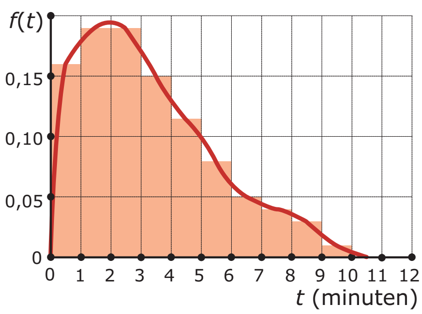
Opmerking: het woord normaal in ‘normale verdeling’ zegt iets over de vorm van de verdeling. De bovenstaande eigenschappen mag je dan gebruiken.



Figuur 5

Voorbeeld 1

In een callcenter is de transactietijd T de tijd die nodig is om een klant te woord te staan. T is een continue kansvariabele.
Waarom zal T niet normaal verdeeld zijn?
Hoe zou je $P(T \leq 4{,}5)$ kunnen bepalen?



Figuur 6

Antwoord

De bijbehorende kansverdeling zal niet symmetrisch zijn: de kans op een kleine transactietijd is veel groter dan die op een grote.

Schat de oppervlakte onder de kromme om $P(T \leq 4{,}5)$ te bepalen, bijvoorbeeld door de staafjes met een klassenbreedte van 1 te gebruiken.
Dan vind je $P(T \leq 4{,}5) \approx 0{,}73$.

Opgave 4

Gebruik de gegevens uit **Voorbeeld 1**.

a Waarom is de transactietijd niet normaal verdeeld?

b Hoe groot is de kans op meer dan tien minuten transactietijd?

c Laat zien, hoe je $P(T \leq 4{,}5)$ kunt bepalen.

Opgave 5

Het lijkt mischien dat in de praktijk bij vrijwel alle continue statistische variabelen een normale verdeling past. Dat is niet het geval: in veel situaties is een verdeling niet symmetrisch. In welke van de volgende gevallen is de verdeling niet symmetrisch?

a het vulgewicht van machinaal gevulde pakken suiker

b de armlengte van volwassen mannen

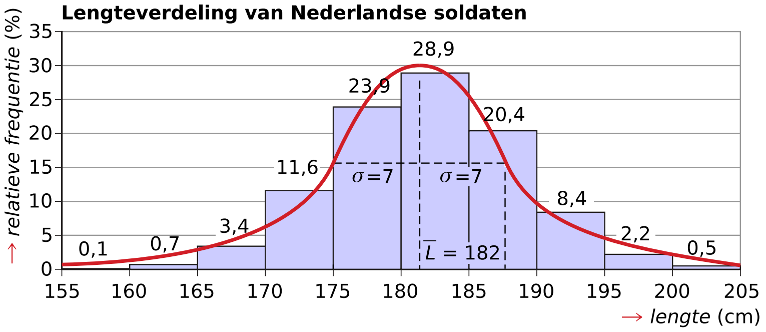
c het gewicht van volwassen mannen

d het inkomen van alle Nederlanders

e de wachttijd bij een helpdesk

Voorbeeld 2

Bekijk de applet



Figuur 7

Bekijk de normaalkromme bij een toevalsvariabele L. Bij elke waarde van L wordt de oppervlakte van het gebied links van die waarde onder de kromme berekend. Voor elke normale toevalsvariabele L geldt bij benadering:

- 68% van de waarden die L kan aannemen ligt tussen $\mu - \sigma$ en $\mu + \sigma$;
- 95% van de waarden die X kan aannemen ligt tussen $\mu - 2\sigma$ en $\mu + 2\sigma$.

Controleer de eerste vuistregel voor de variabele L met $\mu(L) = 182$ en $\sigma(L) = 7$.

Antwoord

Tussen $182 - 7 = 175$ en $182 + 7 = 189$ zitten twee hele staven en $\frac{4}{5}$ van de daaropvolgende staaf.

Dat is $23{,}9 + 28{,}9 + \frac{4}{5} \cdot 20{,}4 = 69{,}12\%$. Dat is ongeveer 68%.

Controleer op dezelfde manier de tweede vuistregel: 95% zit tussen $182 - 14$ en $182 + 14$.

Opgave 6

In **Voorbeeld 2** worden de vuistregels gecontroleerd voor een normale verdeling met $\mu = 182$ en $\sigma = 7$.

a Controleer de tweede vuistregel zelf.

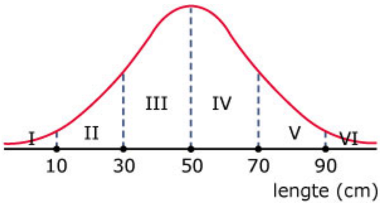
b Hoeveel procent hoort er volgens de vuistregels bij het gebied tussen $\mu - 2\sigma$ en $\mu - \sigma$?

c Teken een normaalkromme met $\mu = 170$ en $\sigma = 10$. Met grenzen $\mu - 2\sigma$, $\mu - \sigma$, μ, $\mu + \sigma$ en $\mu + 2\sigma$ kun je het gebied onder de normaalkromme in zes delen verdelen. Zet in elk van die delen het juiste percentage.

d Hoeveel procent zit er in het gebied tussen $\mu - 3\sigma$ en $\mu + 3\ \sigma$?

Opgave 7

De lengtes van de buxusplanten bij een plantenkweker zijn normaal verdeeld met een gemiddelde van 50 cm en een standaardafwijking van 20 cm. De normaalkromme geeft de lengteverdeling weer. De buxusteler verdeelt de planten in zes categorieën van 20 cm.



Figuur 8

a Bepaal bij benadering hoeveel procent van de planten tot elke categorie behoren.

b Hoeveel procent van de planten is groter dan 70 cm?

c Hoeveel procent van de planten heeft een lengte tussen 30 en 90 cm?

d Hoeveel procent van de planten is minstens 30 cm lang?

Verwerken

Opgave 8

Een supermarkt verkoopt spliterwten in pakken van 500 gram. Veel klanten vermoeden dat in minstens een derde van de pakken te weinig spliterwten zitten. Zij dienen een klacht in bij de directie. Een consumentenorganisatie wordt gevraagd de klacht te onderzoeken. Zij neemt een steekproef van 100 pakken. Het gemiddelde gewicht van de pakken blijkt 502 gram met een standaardafwijking van 8 gram. Verder blijkt het gewicht van de pakken spliterwten normaal verdeeld te zijn.

a Maak een klokvormige kromme bij de verdeling van de pakken spliterwten.

b Ongeveer hoeveel pakken uit de steekproef wijken meer dan één keer de standaardafwijking af van het gemiddelde?

c Ongeveer hoeveel pakken uit de steekproef hebben een gewicht van minder dan 510 gram?

d Kun je bepalen hoeveel procent van de pakken meer weegt dan 511 gram?

e Maak een schatting van het percentage van de pakken dat minder weegt dan 500 gram. Worden de klagers in het gelijk gesteld?

Opgave 9

In een fabriek worden kilopakken suiker machinaal gevuld. Volgens de Europese norm mag niet meer dan 2,5% van de pakken suiker minder dan 1000 gram bevatten, maar deze fabriek lijkt daar moeite mee te hebben.

Bekijk de vulgewichten van dertig pakken suiker uit deze fabriek.

998,3	994,9	1003,5	999,6	1001,0	1001,1
997,9	1003,0	999,5	999,4	991,7	998,0
1001,8	997,8	999,3	995,1	1000,0	998,8
996,3	999,9	999,5	998,6	999,0	1001,5
1000,8	1001,4	998,1	997,5	995,3	998,2

Tabel 3

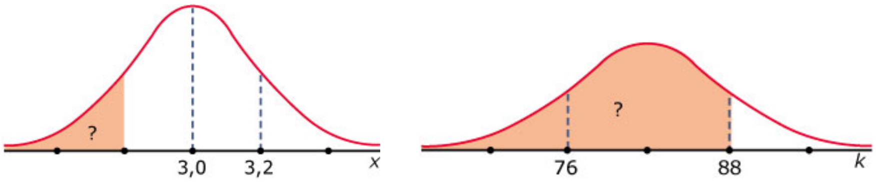
a Hoeveel procent van deze pakken suiker is lichter dan 1000 gram?

b Maak een histogram van de vulgewichten uit de tabel. Gebruik klassen met een klassenbreedte van 1 gram. Laat zien dat de fabrikant met een beetje goede wil kan beargumenteren dat de vulgewichten van deze machine bij benadering een symmetrische klokvormige verdeling hebben.

c Bereken het gemiddelde en de standaardafwijking van deze vulgewichten in één decimaal.

Opgave 10

Bekijk de twee normale verdelingen. Ga er van uit dat de gegeven waarden passen bij de vuistregels.



Figuur 9

a Geef bij elk van deze normaalkrommen de waarden van μ en σ.

b Bepaal het percentage dat hoort bij het aangegeven gebied.

Opgave 11

Een maat voor iemands intelligentie is het intelligentiequotiënt (IQ). Dat is de score op een intelligentietest vergeleken met die van leeftijdsgenoten. Het IQ is normaal verdeeld met een gemiddelde van 100 en een standaardafwijking van 15.

a Hoeveel procent van de mensen heeft een IQ tussen 85 en 115?

b Hoeveel procent van de mensen heeft een IQ van meer dan 130?

c Hoe groot is de kans dat het IQ van een willekeurige voorbijganger minder is dan 130?

d Met welk IQ behoor je tot de mensen die de 16% laagste scores hebben?

Opgave 12

Van twee leeftijdsgroepen zijn de scores voor een test verzameld. De scores van beide groepen zijn bij benadering normaal verdeeld. Bekijk de gemiddelden en de standaardafwijkingen in het overzicht.

	12-jarigen	16-jarigen
aantal tests	500	800
μ	48	56
σ	8	12

Tabel 4

a Hoe groot ongeveer is de kans dat een 12-jarige beter scoorde dan een gemiddelde 16-jarige?

b Schets de twee bijbehorende normaalkrommen langs één horizontale as en geef speciale aandacht aan de tophoogten van de twee krommen. Zijn ze even hoog? Zo nee, welke ligt hoger?

Opgave 13

De firma Sanove fabriceert stukken zeep. De stukken zeep worden machinaal gemaakt. De machine is zo ingesteld dat het gewicht van de stukken zeep normaal verdeeld is met een gemiddelde van 93 gram en een standaardafwijking van 1,4 gram.

Het kan gebeuren dat de machine niet goed functioneert. Dan hebben te veel stukken zeep niet het gewenste gewicht. De afdeling Quality Control (QC) van Sanove gebruikt verschillende manieren om dit te controleren. Eén van deze manieren heet de 2_{2s}-regel.

Bij de 2_{2s}-regel wordt per dag drie keer aselect een stuk zeep gekozen. De machine wordt opnieuw ingesteld zodra die dag twee keer ‘achter elkaar’ het gewicht van een stuk zeep meer dan twee keer de standaardafwijking afwijkt van het gemiddelde. Als van de eerste twee gecontroleerde stukken zeep het gewicht meer dan twee keer de standaardafwijking van het gemiddelde afwijkt, vindt die dag geen derde controle plaats.
Er is een kans dat QC de machine opnieuw laat instellen terwijl de machine in orde is. Hoe groot is deze kans ongeveer?

Toepassen

Opgave 14: Kilopakken suiker vullen

In een fabriek worden kilopakken suiker machinaal gevuld. Volgens de Europese norm mag niet meer dan 2,5% van de pakken suiker minder dan 1000 gram bevatten. Bekijk het bestand **met de vulgewichten van honderd pakken suiker**.

a Maak een histogram van de vulgewichten van machine 1. Gebruik klassen met een klassenbreedte van 1 gram. Laat zien dat de vulgewichten van deze machine bij benadering een symmetrische klokvormige verdeling hebben.

b Bereken het gemiddelde en de standaardafwijking van deze vulgewichten in gehele grammen nauwkeurig.

c Hoeveel procent van de pakken suiker is volgens dit histogram lichter dan 1000 gram?

d Je kunt het histogram benaderen door een normaalkromme met het zojuist berekende gemiddelde en de standaardafwijking.

Hoeveel procent van de pakken is lichter dan 1000 gram volgens deze normale verdeling?

e Hoeveel procent van de vulgewichten wijkt volgens de vuistregels van de normale verdeling minder dan één standaardafwijking van het gemiddelde af? Klopt dat met deze frequentieverdeling?

Opgave 15: Lichaamslengtes van 5001 vrouwen

Open het bestand **Enkele lichaamsafmetingen van 5001 vrouwen uit 1947**. Hierin zie je o.a. een tabel met lichaamslengtes in cm van de 5001 vrouwen uit het onderzoek in 1947 van Freudenthal en Sittig in opdracht van De Bijenkorf.

a Bereken met Excel de gemiddelde lichaamslengte en de standaardafwijking.

b Teken met Excel een histogram van de lichaamslengtes en benader dit met een normaalkromme waarin je gemiddelde en standaardafwijking aangeeft.

Neem nu verder aan dat de lichaamslengte L van vrouwen normaal is verdeeld met de eerder berekende waarden voor het gemiddelde μ en de standaardafwijking σ.

c Is statistische variabele L een discrete of een continue statistische variabele?

Beargumenteer je antwoord.

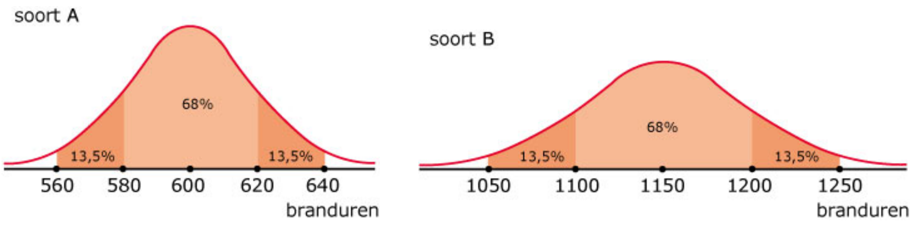
d 95% van de lichaamslengtes zit tussen $\mu - a$ en $\mu + a$. Hoe groot is a?

e Welke minimale lengte hebben de 16% grootste lichaamslengtes?

Testen

Opgave 16

Van twee soorten lampen is de levensduur van 500 exemplaren gemeten. Het aantal branduren blijkt vrijwel normaal verdeeld te zijn. Bekijk de bijpassende normaalkrommen. Enkele percentages zijn gegeven.



Figuur 10

Van soort A is het gemiddelde $\mu_A = 600$ branduren en de standaardafwijking $\sigma_A = 20$ uur.

a Hoeveel procent van de lampen van soort A brandt minder dan 600 uur?

b Hoeveel procent van de lampen van soort A brandt minder dan 620 uur?

c Hoeveel is het gemiddeld aantal branduren van de lampen van soort B? En hoeveel is de standaardafwijking van de lampen van soort B?

Beargumenteer je antwoord.

d Waarom heeft de normale verdeling bij soort B een top die minder hoog is dan die van de normale verdeling van soort A?

e Ongeveer hoeveel procent van de lampen van soort B brandt langer dan 1250 uur?

Opgave 17

Bij een groep van 1000 mannen is de bloeddruk normaal verdeeld met een gemiddelde van 128,5 mm Hg met een standaardafwijking van 12,5 mm Hg.

a Maak een normaalkromme bij de bloeddrukverdeling van deze groep mannen.

b Hoeveel van de mannen hebben een bloeddruk van minder dan 141?

c Hoeveel mannen hebben een bloeddruk die meer dan twee keer de standaardafwijking afwijkt van de gemiddelde bloeddruk?

d Kun je precies bepalen hoeveel procent van de mannen een bloeddruk heeft van meer dan 150?

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt. Klik op in de marge bij de betreffende opgave. Uw mailprogramma wordt dan geopend waarbij het emailadres en onderwerp al zijn ingevuld. U hoeft alleen uw opmerkingen nog maar in te voeren.

Email: `f.spijkers@math4all.nl`

Met de Math4All Foliostraat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij `a.f.otten@math4all.nl` een gratis inlog voor de maatwerkdienst aanvragen.