3.2 Vuistregels

Inleiding

Je hebt al gezien dat in de statistiek de normale verdeling een belangrijke rol speelt. Onder andere bij het bepalen van betrouwbaarheidsintervallen en foutenmarges bij het schatten van populatieproporties en populatiegemiddelden. Daarbij maak je gebruik van een aantal vuistregels die voor elke normale verdeling gelden.

Je leert in dit onderwerp

- werken met drie vuistregels voor de normale verdeling.
- dat een steekproevenverdeling bij benadering een normale verdeling is.

Voorkennis

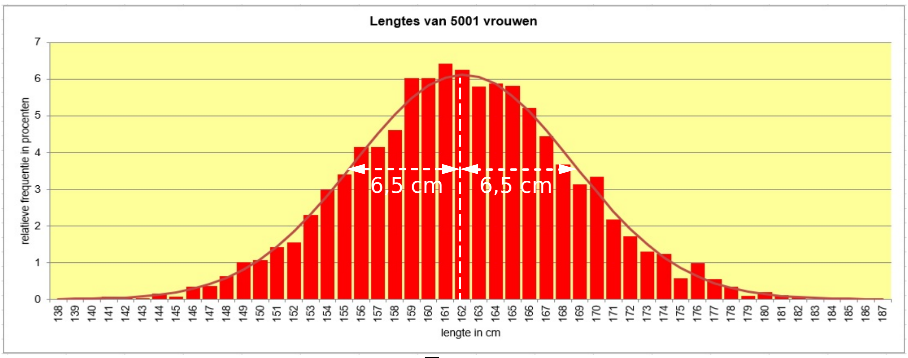
- soorten statistische variabelen herkennen;
- de begrippen onderzoek, steekproef, populatie en representatief, simulatie;
- het begrip normale verdeling met gemiddelde en standaardafwijking.

Verkennen

Opgave V1

Je ziet een staafdiagram van de lengtes van 5001 vrouwen uit de dataset **Statistiek Bijenkorf 1947**.

Die verdeling heeft een vrijwel zuivere ‘klokvorm’. Je noemt dit een normale verdeling, elke normale verdeling heeft zo'n klokvorm. De gemiddelde lengte $\overline{L} \approx 162$ cm en de standaardafwijking $\sigma(L) \approx 6{,}5$ cm zijn in de figuur aangegeven.



$\overline{L} \approx 162$ $\quad$ $s_L \approx 6{,}5$

Figuur 1

Beide getallen leggen de normale verdeling volledig vast, ze zijn karakteristiek voor deze normale verdeling. De standaardafwijking geef je aan met een Griekse letter s, de ‘sigma’ in $\sigma(L)$. Voor het gemiddelde gebruik je ook wel een Griekse letter m, de ‘mu’ in $\mu(L)$. Op grond van alleen het gemiddelde en de standaardafwijking kan elke normale verdeling worden getekend.

a Welke lengtes horen bij het interval $\mu - 3\sigma, \mu + 3\sigma$?

b Ga na dat bijna 100% van de lengtes van deze vrouwen binnen dat interval vallen.

c Hoeveel procent van de lengtes valt binnen het interval $[\mu - \sigma, \mu + \sigma]$?

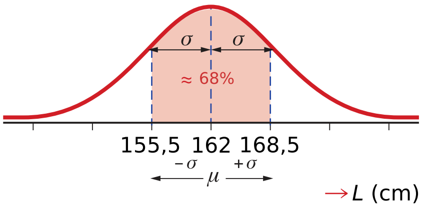
Uitleg 1

Bekijk de applet.

Bij statistisch onderzoek maak je vaak gebruik van de normale verdeling. Een goed voorbeeld komt uit een onderzoek uit 1947 van de Bijenkorf naar de lengtes van Nederlandse vrouwen. Die lengtes waren ongeveer normaal verdeeld met een gemiddelde van $\mu(L) \approx 162$ cm en een standaardafwijking van $\sigma(L) \approx 6{,}5$ cm.

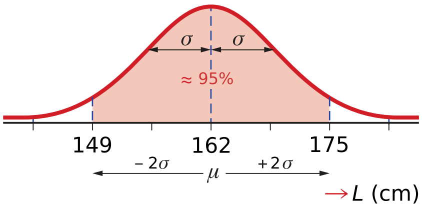
Bekijk de normale verdeling van de lichaamslengtes van de vrouwen.

Bekijk de lichaamslengtes tussen $162 - 6{,}5 = 155{,}5$ cm en $162 + 6{,}5 = 168{,}5$ cm.



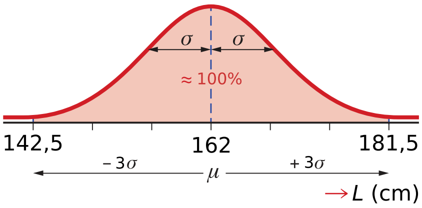
Figuur 2

Bekijk de lichaamslengtes tussen $162 - 2 \cdot 6{,}5 = 149$ cm en $162 + 2 \cdot 6{,}5 = 175$ cm.



Figuur 3

Bekijk de lichaamslengtes tussen $162 - 3 \cdot 6{,}5 = 142{,}5$ cm en $162 + 3 \cdot 6{,}5 = 181{,}5$ cm.



Figuur 4

Experimenteer zelf nog met andere lichaamslengtes.

Omdat de vorm van een normale verdeling altijd hetzelfde is zijn ook de percentages van elke normale verdeling hetzelfde. Als $\mu(L)$ en $\sigma(L)$ bekend zijn, ligt de normale verdeling helemaal vast. Je kunt percentages aflezen of berekenen. Het percentage onder elke normale verdeling met waarden

- tussen $\mu - \sigma$ en $\mu + \sigma$ cm is 68%;
- tussen $\mu - 2 \cdot \sigma$ en $\mu + 2 \cdot \sigma$ cm is 95% ;
- tussen $\mu - 3 \cdot \sigma$ en $\mu + 3 \cdot \sigma$ cm is nagenoeg 100%.

Deze percentages zijn vuistregels, want ze zijn afgerond. Deze vuistregels gebruik je vaak bij berekeningen.

Opgave 1

Gegeven is de normale verdeling uit **Uitleg 1**. Gebruik de vuistregels.

a De Bijenkorf maakt kleding voor deze groep vrouwen. Hoeveel procent van de kleding moet er gemaakt worden voor de vrouwen met een lichaamslengte tussen 155,5 en 168,5 cm?

b Hoeveel procent van de kleding moet er gemaakt worden voor de vrouwen met een lichaamslengte tussen 149 en 175 cm?

c Hoeveel procent van deze kleding moet er gemaakt worden voor de vrouwen met een lichaamslengte tussen 162 en 168,5 cm?

Opgave 2

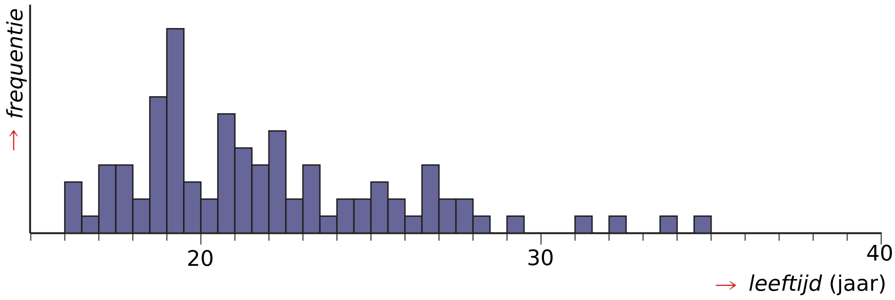
Het gewicht van potten jam is normaal verdeeld. Het gemiddelde is 200 gram en de standaardafwijking is 3 gram. Er mag maximaal 3% van de potten jam minder dan 194 gram wegen. Is aan deze eis voldaan?

Uitleg 2

Er is een heel groot concert met tienduizenden bezoekers. De organisatoren van het concert willen de gemiddelde leeftijd van de bezoekers weten.

Bij elk van de 50 ingangen zetten ze een enquêteur die aan elke volgende 100e bezoeker de leeftijd vraagt. Zo worden er 50 steekproeven genomen. Ga ervan uit dat deze steekproeven representatief zijn. Omdat niet iedereen wordt ondervraagd, kun je de gemiddelde leeftijd niet precies te weten komen. Je kunt deze alleen maar schatten.

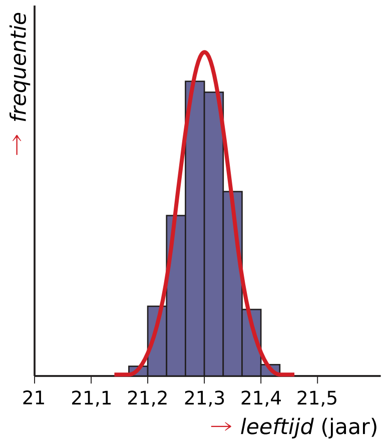
In het linker histogram zie je de resultaten van één van de 50 steekproeven.



Figuur 5

Het lijkt erop dat de leeftijden van de concertbezoekers niet een normale verdeling hebben, want de verdeling ziet er niet symmetrisch uit.
Van alle 50 steekproeven die genomen zijn is de gemiddelde leeftijd uitgerekend. Ook deze gegevens zijn in een histogram verwerkt.

Dit histogram heeft ongeveer een klokvorm en is daarom te benaderen met een normale verdeling. De verdeling van de gemiddelden van veel, minstens 50 steekproeven, is altijd bij benadering een normale verdeling.



Figuur 6

Opgave 3

Bestudeer **Uitleg 2**. Bekijk de verdeling van de gemiddelden van de leeftijd in de steekproeven.

a Lees uit de figuur het gemiddelde van deze normale verdeling af.

b Geef met de figuur een schatting van de standaardafwijking van deze normale verdeling.

c Stel je neemt opnieuw een serie van die steekproeven. Tussen welke leeftijden zal het gemiddelde in 95% van deze steekproeven liggen?

Opgave 4

Bestudeer **Uitleg 2**. De verkopers van drinken willen nog nauwkeuriger weten hoe de leeftijden van concertgangers verdeeld zijn. Noem twee mogelijke manieren om dit te bereiken.

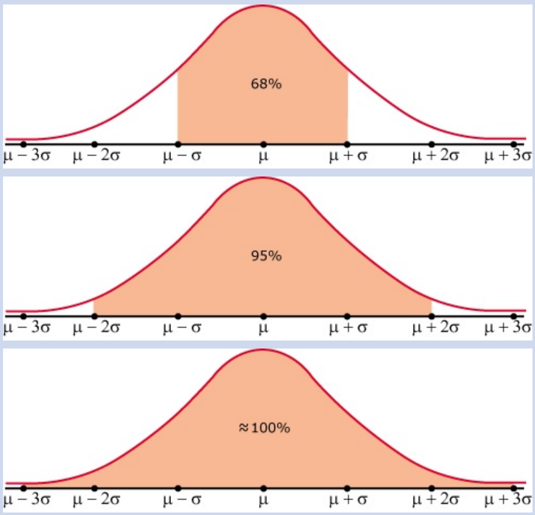
Theorie en voorbeelden

Om te onthouden

Deze **vuistregels** gelden voor alle normaal verdeelde variabelen X:

- Ongeveer 68% van alle waarden van X ligt tussen $\mu - \sigma$ en $\mu + \sigma$. De buigpunten van de normale verdeling liggen ook op deze afstand van μ.
- Ongeveer 95% van alle waarden van X ligt tussen $\mu - 2 \cdot \sigma$ en $\mu + 2 \cdot \sigma$.
- Nagenoeg 100% van alle waarden van X ligt tussen $\mu - 3 \cdot \sigma$ en $\mu + 3 \cdot \sigma$.

Er zijn veel statistische variabelen die niet normaal verdeeld zijn. De gegevens van een representatieve steekproef van zo'n populatie zijn dan ook NIET normaal verdeeld, want een representatieve steekproef heeft dezelfde verdeling als de populatie.



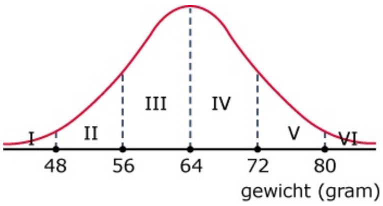
Figuur 7

Van elke steekproef kun je het gemiddelde of het gemiddelde percentage uitrekenen. Er blijkt iets bijzonders te gebeuren als je dit met veel steekproeven doet: de verdeling van de gemiddeldes of de gemiddelde percentages van de steekproeven lijkt steeds meer op een normale verdeling als het aantal steekproeven groter wordt. Dit heet een **steekproevenverdeling**. Bekijk de applet in het **Practicum**.

In de praktijk zijn minstens 50 steekproeven nodig voordat de steekproevenverdeling op een normale verdeling gaat lijken.

Voorbeeld 1

Eieren kunnen op grond van hun gewicht in klassen verdeeld worden. De gewichten zijn vrijwel normaal verdeeld met een gemiddelde van 64 gram en een standaardafwijking van 8 gram. In deze figuur zie je zes gewichtsklassen voor de gewichten van de eieren. Geef bij elke klasse de grenzen aan van het gewicht van de eieren die er in zitten en bepaal hoeveel procent van de eieren het betreft.



Figuur 8

Antwoord

De klassengrenzen zijn zo gekozen dat ze precies passen bij de vuistregels voor de normale verdeling:

- Klasse I: ≤ 48 gram met 2,5% van de eieren.
- Klasse II: tussen 48 en 56 gram ligt 13,5% van de eieren.
- Klasse III: tussen 56 en 64 gram ligt 34% van de eieren.
- Klasse IV: tussen 64 en 72 gram ligt 34% van de eieren.
- Klasse V: tussen 72 en 80 gram ligt 13,5% van de eieren.
- Klasse VI: ≥ 80 gram met 2,5% van de eieren.

Opgave 5

Bekijk **Voorbeeld 1**.

a Leg uit waaraan je ziet dat de klassengrenzen precies passen bij de vuistregels.

b Hoeveel procent van de eieren is lichter dan 80 gram?

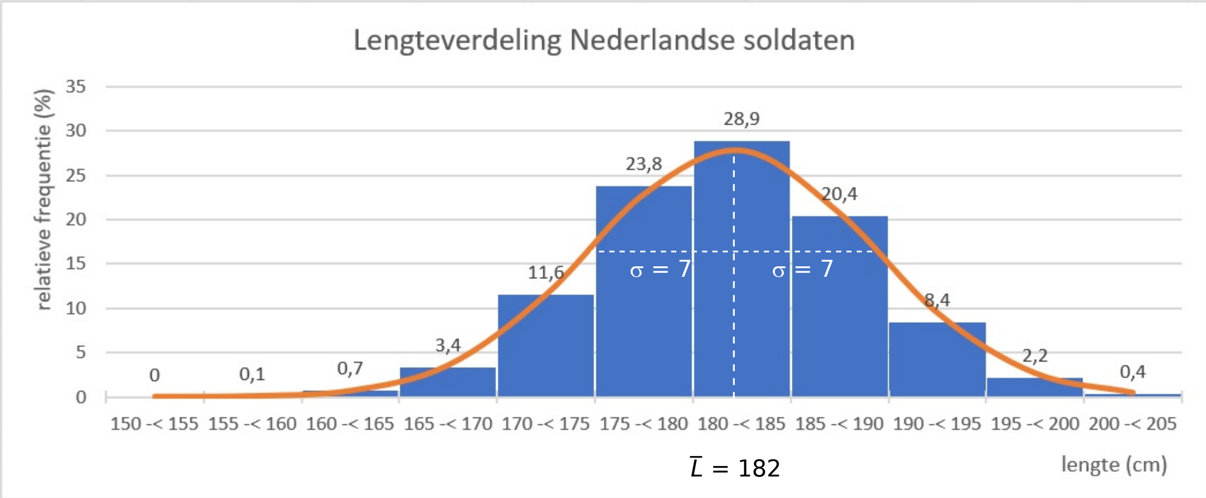
c Hoeveel procent van de eieren is zwaarder dan 72 gram?

d Je koopt bij de eierboer 100 ongesorteerde eieren. Hoeveel daarvan zullen waarschijnlijk minder dan 56 gram wegen?

e Waarom kun je niet met zekerheid zeggen dat er 16 eieren minder dan 56 gram wegen?

Voorbeeld 2

Bekijk de applet: Normale verdeling



Figuur 9

Hier zie je een normale verdeling van de lengte van Nederlandse soldaten met $\mu = 182$ en $\sigma = 7$.

Bereken het percentage van deze soldaten met een lengte van meer dan 189 cm.

Antwoord

De eerste vuistregel zegt dat 68% van de soldaten een lengte heeft in het interval $[\mu - \sigma, \mu + \sigma] = [175,189]$.

Dit betekent dat een percentage van $100 - 68 = 32\%$ daar buiten ligt.

Boven de 189 cm zit dus 16% van deze soldaten.

Opgave 6

Bekijk **Voorbeeld 2**.

a Tussen welke lengtes zit 95% van deze soldaten?

b Je wilt weten hoeveel procent van de soldaten langer is dan 195 cm. Waarom kun je dit niet met de vuistregels beantwoorden?

c Bepaal het percentage van de soldaten dat langer is dan 196 cm.

d Tussen welke lengtes zitten deze soldaten vrijwel allemaal?

Opgave 7

De gemiddelde lengte van vrouwen is bij benadering normaal verdeeld. In 2010 was de gemiddelde lengte van de vrouwen in Nederland 170 cm met een standaardafwijking van 6,5 cm.

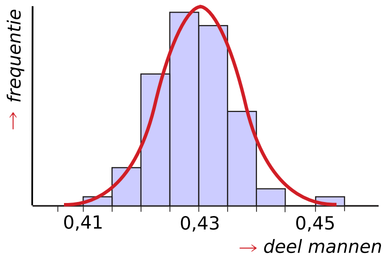
a Schets hierbij een normale verdeling met de grenzen die horen bij de vuistregels.

b Schrijf drie uitspraken op over de lengte van vrouwen in 2010 gebaseerd op de vuistregels.

Voorbeeld 3

Er is een heel groot concert. Er zijn 50 ingangen. De verkopers van shirts willen weten hoe groot het deel mannen er is. Bij elke ingang zetten ze een enquêteur die van elke tiende bezoeker opschrijft of het een man is. Bij elkaar hebben de verkopers dus 50 steekproeven genomen. We nemen aan dat deze steekproeven representatief zijn.

Van de 50 steekproeven wordt uitgerekend welk deel van de bezoekers man is. Daar komt een getal tussen 0 (geen mannen) en 1 (allemaal mannen) uit. Deze 50 getallen worden in een histogram gezet. Het histogram lijkt op een normale verdeling. Deze normale verdeling is getekend.



Figuur 10

Deze normale steekproevenverdeling heeft een gemiddelde $\mu = 0{,}43$. Voor de standaardafwijking geldt $\sigma = 0{,}01$.

Waarom zal in 95% van de steekproeven het deel van de bezoekers dat man is, tussen de 0,41 en 0,45 liggen?

Antwoord

De steekproeven zijn representatief, dus er is alleen sprake van toevallige verschillen. Het deel mannen in deze steekproeven is normaal verdeeld met $\mu = 0{,}43$. De getallen 0,41 en 0,45 liggen daar $2 \cdot \sigma$ vanaf. Volgens de vuistregels is het percentage van de steekproeven waarin het deel mannen bij het concert tussen 0,41 en 0,45 ligt, dus 95%.

Opgave 8

Bekijk **Voorbeeld 3**.

a Bij hoeveel procent van de steekproeven ligt het deel van de bezoekers dat man is, tussen de 0,42 en 0,44?

b Bij hoeveel procent van de steekproeven ligt het deel van de bezoekers dat man is, boven de 0,46?

Verwerken

Opgave 9

Een medicijnfabrikant heeft een statistisch onderzoek gedaan naar het gewicht X van tabletten. Het histogram was ongeveer klokvormig met $\mu(X) = 500$ mg en $\sigma(X) = 10$ mg.

a Maak een schets van de bijbehorende normale verdeling. Zet de variabele, de symbolen voor gemiddelde en standaardafwijking, de waarden en de eenheden er in.

b Hoeveel procent van de tabletten is lichter dan het gemiddelde gewicht?

c Geef een schatting van het percentage tabletten dat tussen 490 mg en 510 mg weegt.

Opgave 10

Een maat voor iemands intelligentie is het intelligentiequotiënt, IQ. Het IQ wordt bepaald door de score op een intelligentietest te vergelijken met de gemiddelde score op dezelfde test van leeftijdsgenoten. Het IQ is normaal verdeeld met een gemiddelde van 100 en een standaardafwijking van 15. Maak gebruik van de vuistregels en rond antwoorden af op halve procenten.

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

b Hoeveel procent van de mensen heeft een IQ hoger dan 115?

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

d Hoeveel procent is de kans dat het IQ van een willekeurige voorbijganger minder is dan 130?

Opgave 11

Het gemiddelde gewicht van mannen ouder dan 20 jaar, is bij benadering normaal verdeeld. In 2011 was het gemiddelde gewicht van een man 84 kg. De standaardafwijking bij mannen is 11 kg.

Maak gebruik van de vuistregels en rond antwoorden af op halve procenten.

a Hoeveel procent van de mannen is naar schatting zwaarder dan 73 kg én lichter dan 84 kg?

b Hoeveel procent van de mannen is naar schatting zwaarder dan 73 kg én lichter dan 95 kg?

c Hoeveel procent van de mannen is naar schatting zwaarder dan 73 kg én lichter dan 106 kg?

d Mannen waarvan het gewicht meer dan twee keer de standaardafwijking afwijkt van het gemiddelde, zijn te licht of te zwaar. Hoeveel procent van de mannen is te licht of te zwaar?

(bron: CBS)

Opgave 12

In een groot stadion zijn bezoekers willekeurig door 75 ingangen naar de verschillende vakken in het stadion gegaan. Bij alle ingangen is door middel van een steekproef de leeftijden van de bezoekers gevraagd. Hierdoor zijn 75 steekproeven genomen. Je mag er vanuit gaan dat de steekproeven representatief en voldoende groot zijn. Uit dit onderzoek kwam naar voren dat de leeftijd van de bezoekers niet normaal verdeeld is.

a Leg uit of de leeftijd van de groep bezoekers van één ingang normaal verdeeld is.

b Leg uit of het steekproevengemiddelde van de leeftijd normaal verdeeld is.

Opgave 13

De gemiddelde lengte van mannen is bij benadering normaal verdeeld. In 2010 was de gemiddelde lengte van de mannen in Nederland 181 cm met een standaardafwijking van 7 cm.

a Schrijf de drie vuistregels op voor de lengte van mannen in 2010.

b Schrijf nog minimaal twee uitspraken op over de lengte van mannen in 2010, gebaseerd op de vuistregels.

c In 1981 waren mannen gemiddeld 3 cm kleiner. De standaardafwijking was vrijwel hetzelfde, dus 7 cm. Welke getallen in de antwoorden van de vorige twee vragen veranderen wel, en hoe, en welke niet?

Opgave 14

Er is onderzoek gedaan onder treinreizigers op het traject Den Haag CS - Utrecht CS. Er is gevraagd naar de totale reisduur van vertreklocatie (bv. huis) naar bestemming (bv. werk). De totale reisduur bleek normaal verdeeld te zijn. De standaardafwijking van deze totale reisduur bleek 15 minuten te zijn.

Bekijk nu niet de 2,5% reizigers die de langste totale reisduur hadden. Bekijk ook niet de 2,5% reizigers die de kortste totale reisduur hadden. Wat is het verschil in reistijd tussen langste en kortste totale reisduur van de overgebleven reizigers?

Toepassen

Opgave 15: Kniehoogtes

Open **het bestand met kniehoogtes** in Excel.

Dit is een Excel-bestand met een tabel van de kniehoogtes in cm van de 5001 vrouwen uit het onderzoek dat in 1947 door Freudenthal en Sittig in opdracht van De Bijenkorf is gedaan.

a Bereken met behulp van Excel de gemiddelde kniehoogte en de standaardafwijking.

b Hoe kun je onderzoeken of de kniehoogte normaal verdeeld is?

c Voer het onderzoek bedoeld bij b uit.

Testen

Opgave 16

De gemiddelde lengte van vrouwen is bij benadering normaal verdeeld. In 2010 was de gemiddelde lengte van de vrouwen in Nederland 167,5 cm met een standaardafwijking van 6,5 cm. Maak gebruik van de vuistregels en rond antwoorden af op halve procenten.

a Hoeveel procent van de vrouwen was waarschijnlijk kleiner dan 154,5 cm?

b Hoeveel procent van de vrouwen was waarschijnlijk kleiner dan 180,5 cm?

c Hoe groot is de kans dat een willekeurige vrouw groter is dan 180,5 cm?

Opgave 17

In een concertzaal zijn bezoekers willekeurig door vier ingangen naar binnen gegaan. Bij alle ingangen zijn door middel van een steekproef de leeftijden van de bezoekers gevraagd. Hierdoor zijn vier steekproeven genomen. Je mag ervan uitgaan dat de steekproeven representatief en voldoende groot zijn. Uit dit onderzoek kwam naar voren dat de leeftijd van de bezoekers geen normale verdeling heeft.

a Leg uit of de leeftijd van de groep bezoekers van één ingang een normale verdeling heeft.

b Leg uit of het steekproefgemiddelde van de leeftijd een normale verdeling heeft.

Practicum

Figuur 11

Het **trekken van een aselecte steekproef** kun je simuleren met deze VUstat-apps.
In de eerste simulatie kun je zien welk percentage een aselecte steekproef van bijvoorbeeld 200 (in te stellen bij ‘omvang’) uit een populatie van 5000 oplevert als je het percentage in de steekproef weet (instellen bij ‘percentage Groen’). Je kunt ook heel veel van die steekproeven laten doen en zien hoezeer de percentages gespreid liggen.
In de tweede simulatie trek je een groot aantal steekproeven (steekproefgrootte in te stellen) uit een populatie waarvan het populatiegemiddelde is in te stellen. Het doel is om te laten zien dat die steekproefgemiddelden normaal verdeeld liggen.

- **Simulatie steekproeven uit ja/nee populatie**
- **Simulatie steekproeven om een populatiegemiddelde te schatten**