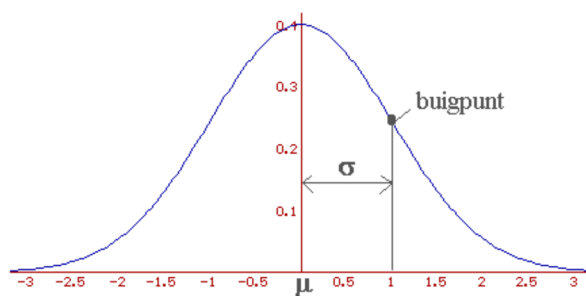


Statistiek en beslissingen



4 HAVO wiskunde A getal en ruimte deel 2

 **hoofdstuk 6 PDF**

0. voorkennis

Centrum- en spreidingsmaten

Centrummaten:

- ✓ het rekenkundig gemiddelde
- ✓ de mediaan
- ✓ de modus

Spreidingsmaten:

- ✓ spreidingsbreedte
- ✓ kwartielafstand
- ✓ standaardafwijking

Zie **verwerken van data**



Steekproeven

Een steekproef is **representatief** als zij een juiste afspiegeling is van de gehele populatie.

Een representatieve steekproef zal in elk geval **voldoende groot** en **aselect** moeten zijn.

De **populatieproportie** p van een populatie is het aantal elementen met een bepaald kenmerk in de populatie gedeeld door totaal aantal elementen in de populatie.

De **steekproefproportie** \hat{p} van een steekproef is het aantal elementen met een bepaald kenmerk in de steekproef gedeeld door het aantal elementen in de steekproef.

Aselect betekent dat elk element van de populatie dezelfde kans heeft om in de steekproef te worden opgenomen.

Opdracht 1

Hier zie je een steelbladdiagram met de proefwerkcijfers van 3B.

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

steel | blad

- Bereken het gemiddelde en de standaardafwijking van de proefwerkcijfers op 1 decimaal nauwkeurig.
- Bepaal de kwartielen en de mediaan en teken het boxplot.
- Bereken de spreidingsbreedte en de kwartielafstand.
- Wat is hier de modus?

Opdracht 2

Van de 1415 leerlingen van een school hebben 380 leerlingen een bijbaantje. Bij een onderzoek naar vrijetijdsbestedingen worden 65 leerlingen ondervraagd. Hiervan blijken 25 leerlingen een bijbaantje te hebben.

- Is het waarschijnlijk dat hier sprake is van representatieve steekproef?
- Hoeveel leerlingen met een bijbaantje zitten er in de steekproef van lengte 65 in het geval \hat{p} gelijk is aan p ?

Opdracht 3

Leerlingen van schoolkrant vragen aan het begin van de kleine pauze aan de leerlingen bij de buitendeur wat ze vinden van het voorgenomen rookverbod op het terrein van school.

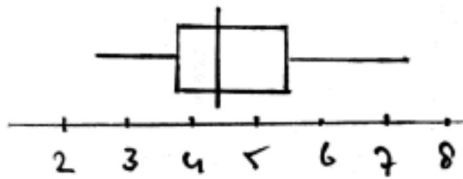
- ✓ Wat is er mis met deze manier van onderzoek?

Opdracht 1

a. $\mu = 4,4$ en $\sigma = 1,2$

b. mediaan = 4,4

$q_1 = 3,75$ en $q_3 = 5,3$



c. spreidingsbreedte = $7,2 - 2,4 = 4,8$

kwartielafstand = $5,3 - 3,75 = 1,55$

d. modus = 4,5

Opdracht 2

a. $p = \frac{380}{1415} \approx 0,27$

$\hat{p} = \frac{25}{65} \approx 0,38$

\hat{p} wijkt erg ver af van p , dus het is niet waarschijnlijk dat hier sprake is van een representatieve steekproef.

b. Voor a als aantal leerlingen met een bijbaantje in de steekproef zou moeten gelden dat $\frac{a}{65} = 0,27$, dus $a \approx 18$ moeten zijn.

Opdracht 3

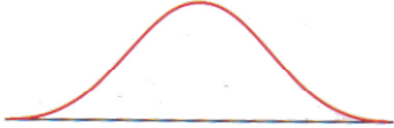
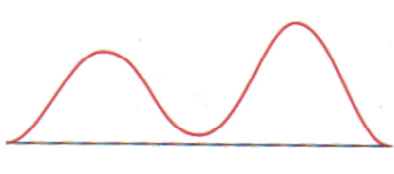
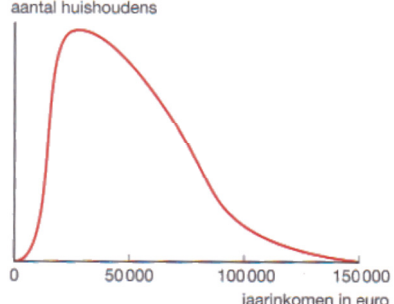
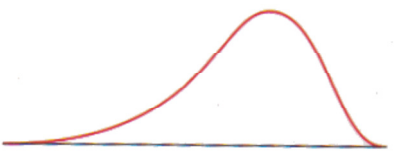
De groep leerlingen die in de kleine pauze naar buiten gaat is geen goede afspiegeling van de schoolpopulatie. Sterker nog de kans dat een 'roker' in de steekproef wordt opgenomen is waarschijnlijk veel groter dan van een 'niet-roker'.



1. soorten verdelingen

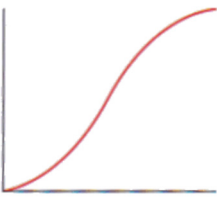

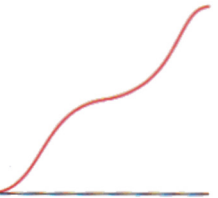

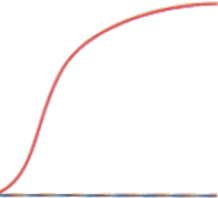

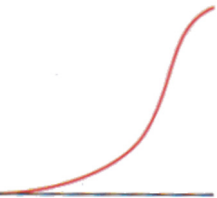

Verdelingen, centrummaten en spreiding

✓ Voorbeelden van verdelingskrommen.

I.	symmetrische verdeling ✓ het gemiddelde, de modus en de mediaan vallen samen	
II.	tweetoppige verdeling ✓ in de figuur hiernaast liggen het gemiddelde en de mediaan tussen te twee toppen ✓ de standaardafwijking is relatief groot	
III.	rechts-scheve verdeling ✓ het gemiddelde en de mediaan liggen rechts van de top ✓ de staart rechts zorgt voor een relatief grote standaardafwijking	
IV.	links-scheve verdeling ✓ het gemiddelde en de mediaan liggen links van de top ✓ de staart links zorgt voor een relatief grote standaardafwijking	

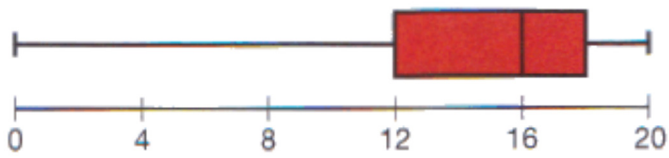
Cumulatieve verdelingskrommen en boxplots

Bij de verdelingen I, II, III en IV hierboven zijn cumulatieve verdelingskrommen en boxplots te schetsen. Omgekeerd kan je uit een cumulatieve verdelingskromme of een boxplot de verdeling herkennen.

 <p>I</p> 	 <p>II</p> 	 <p>III</p> 	 <p>IV</p> 
--	---	--	---

Opdracht 1 (11)

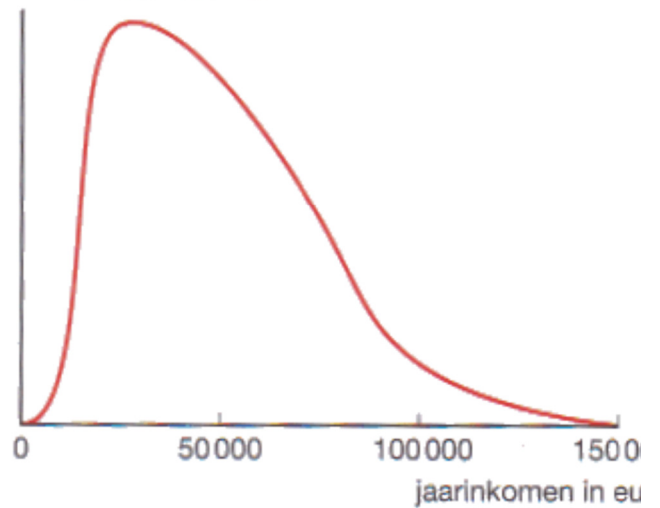
Gegeven de boxplot::



- Schets de bijbehorende verdelingskromme.
- Met wat voor een soort verdeling heb je te maken?
- Hoe ligt het gemiddelde ten opzichte van de mediaan?

Opdracht 2 (A12)

aantal huishoudens

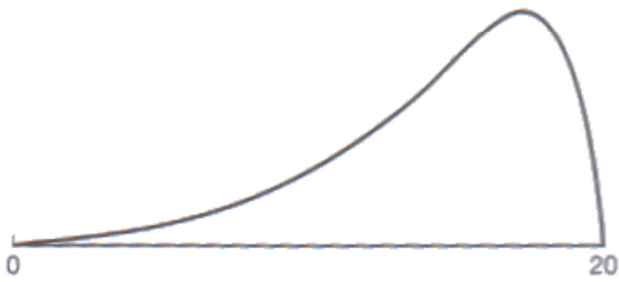


De verdelingskromme geeft informatie over de verdeling van de jaarinkomens van huishoudens een land.

- Hoe ligt de mediaan ten opzichte van de modus bij deze verdeling?
- Schets de relatieve cumulatieve verdelingskromme.
- Schets de boxplot.

Opdracht 1

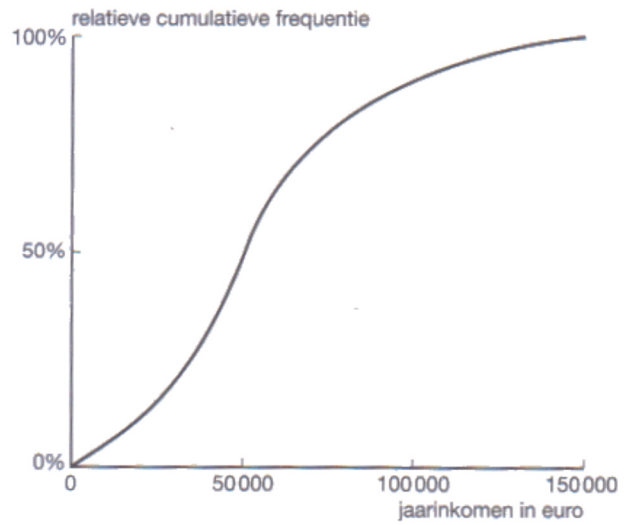
- a. Weinig waarnemingsgetallen aan het begin en veel op het eind



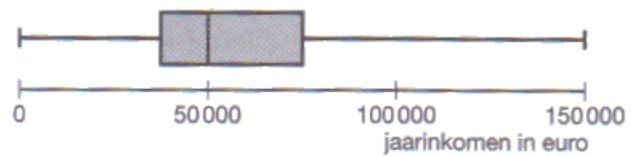
- b. Het is een links-scheve verdeling
c. Het gemiddelde ligt links van de mediaan

Opdracht 2

- a. De mediaan ligt rechts van de modus
b. Het steilste stuk zit ongeveer bij 25.000 euro. De mediaan is ongeveer 50.000

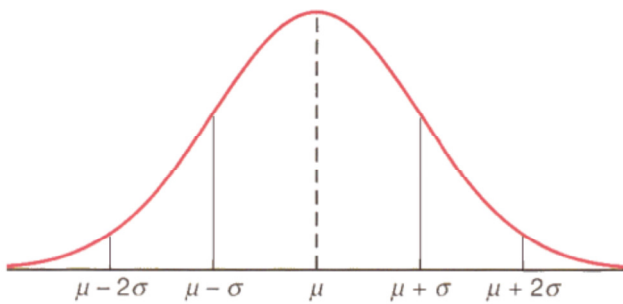


c.



2. de normale verdeling

De normale verdeling



figuur 7.21

Een paar eigenschappen van een normale verdeling:

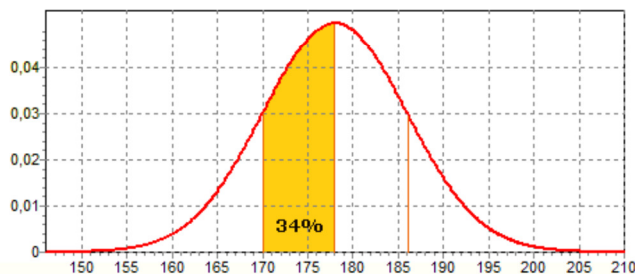
- ✓ **klokvormig**
- ✓ **symmetrisch** t.o.v. het gemiddelde.
- ✓ gemiddelde, mediaan en modus vallen samen
- ✓ de verdeling wordt bepaald door de **verwachtingswaarde** en de **standaarddeviatie**

Vuistregels bij de normale verdeling

- I. **68%** van de gegevens wijkt op z'n hoogst één keer de standaarddeviatie af van de verwachtingswaarde
- II. **95%** van wijkt op z'n hoogst twee keer de standaarddeviatie af van de verwachtingswaarde

Toepassen van de vuistregels

Gegeven is dat de lengte van mannen normaal verdeeld is met $\mu = 178 \text{ cm}$ en $\sigma = 8 \text{ cm}$. Je kunt dan (bijvoorbeeld) de volgende verdeling maken:

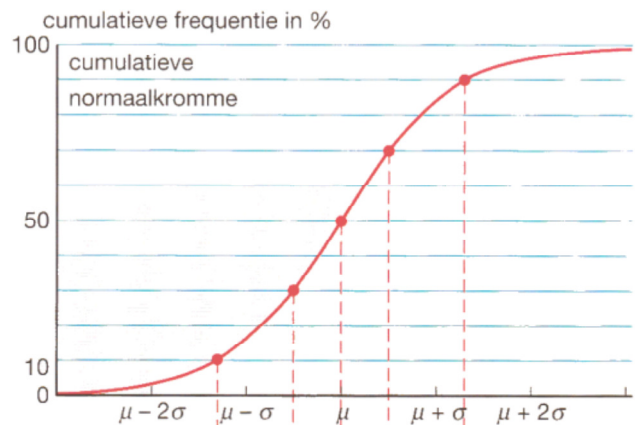


Dus 34% van de mannen heeft een lengte tussen 170 en 178 cm.

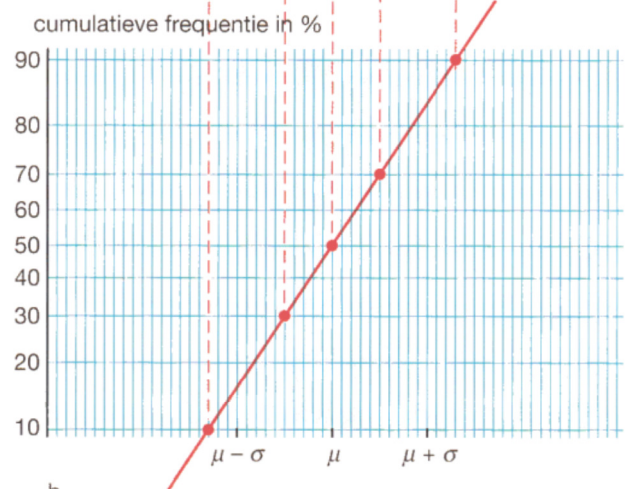
Normaal waarschijnlijkheidspapier

Bij een normale verdeling hoort een rechte lijn op **normaal-waarschijnlijkheidspapier**.

Je kunt μ aflezen bij de relatieve cumulatieve frequentie 50. Je kunt $\mu + \sigma$ aflezen bij de relatieve cumulatieve frequentie 84. Hieruit volgt σ .



a



b

figuur 7.22

Opdracht 1 (18)



Het gewicht van de mandarijnen uit een grote partij is normaal verdeeld met een gemiddelde van 80 gram. Verder is bekend dat 16% van de mandarijnen minder dan 76 gram weegt.

- ✓ Bereken de standaardafwijking

Opdracht 2 (A19)



Van 200 konijnen is het gewicht normaal verdeeld met een gemiddelde van 2,1 kg en een standaardafwijking van 0,3 kg.

- Hoeveel procent van de konijnen is zwaarder dan 2,7 kg?
- Hoeveel konijnen hebben een gewicht tussen 1,5 en 2,4 kg?
- Hoeveel konijnen zijn er lichter dan 1,8 kg?
- Wat weet je van de gewichten van de zwaarste konijnen in deze groep?

Opdracht 3 (A25)

Van een grote groep mannen is de lengte normaal verdeeld. Verder is bekend dat 15% korter is dan 1,70 m en 25% langer dan 1,85 m. Gebruik het normaal-waarschijnlijkheidspapier om uit te zoeken hoe groot het gemiddelde en de standaarddeviatie van de lengte zijn. Rond af op gehele centimeters.

- ✓  normaal-waarschijnlijkheidspapier

Opdracht 4 (A26)

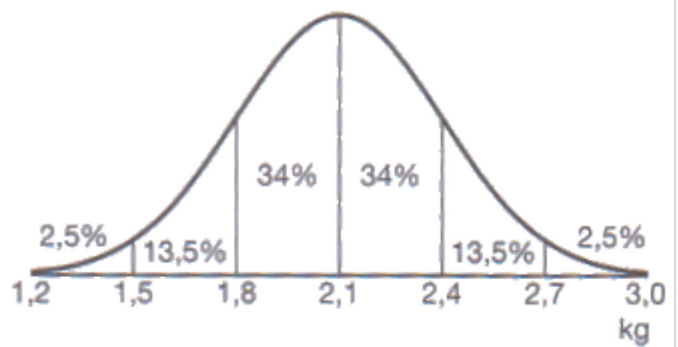
Een bioloog onderzoekt van enkele soorten planten de lengte van de bladeren. Van elke soort zet hij de resultaten in mm uit op normaal-waarschijnlijkheidspapier. Telkens blijkt de lengte normaal verdeeld te zijn.

- De lijnen van de soorten A en B zijn evenwijdig. Wat weet je van de bladlengte van de soorten A en B ?
- De lijn bij soort A is steiler dan die hoort bij soort C . Wat volgt hieruit?
- De lijnen bij de soorten C en D snijden elkaar is het punt $(48,80)$. Wat kan je zeggen van de soorten C en D ?
- De gemiddelde bladlengte van soort B is gelijk aan die van soort D . Wat weet je van de lijnen bij de soorten B en D op normaal-waarschijnlijkheidspapier?

Opdracht 1

16% weegt minder dan 76 gram, dus 76 ligt één standaardafwijking van het gemiddelde af. De standaardafwijking is $80-76=4$ gram.

Opdracht 2



- 2,5%
- $13,5\% + 68\% = 81,5\%$. $0,815 \cdot 200 = 163$ konijnen
- $2,5\% + 13,5\% = 16\%$, dus $0,16 \cdot 200 = 32$ konijnen
- $\frac{5}{200} \cdot 100\% = 2,5\%$, dus deze hebben een gewicht van meer dan 2,7 kg.

Opdracht 3

Bij 170 cm hoort 15%.

Bij 185 cm hoort 75%.

Lees af bij 50% dat $\mu = 179$ cm.

Lees af bij 84% dat $\mu + \sigma = 188$, dus

$$\sigma = 188 - 179 = 9 \text{ cm}$$

✓  **normaal-
waarschijnlijkheidspapier**

Opdracht 4

- Evenwijdig betekent dezelfde standaardafwijking
- Dat de bladlengte van soort A een kleinere standaardafwijking heeft dan die van soort C.
- Dat zowel bij soort C als bij soort D 80% van de bladeren korter is dan 45 cm.
- De lijnen bij B en D moeten elkaar dan snijden op een hoogte van 50.

3. betrouwbaarheidsintervallen

Gemiddelde en standaardafwijking van een proportie

Bij een steekproevenverdeling is de steekproefproportie \hat{p} bij een voldoende grote steekproef n bij benadering normaal verdeeld.

Hierbij is het gemiddelde μ gelijk aan de populatieproportie.

Voor de standaardafwijking σ geldt:

$$\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Voorbeeld

Men neemt 800 steekproeven van lengte 250 bij een populatieproportie van 0,350. Hieruit volgt dat de bijbehorende standaardafwijking gelijk is aan 0,030.

$n = 250$ en $\hat{p} = p = 0,350$ geeft:

$$\sigma = \sqrt{\frac{0,350 \cdot 0,650}{250}} = 0,0301\dots \approx 0,030$$

✓ Zie opdracht 1

Berekening van de steekproefomvang

Je kunt bij een gegeven betrouwbaarheidsinterval de steekproefomvang berekenen.

Aanpak

- ✓ Bepaal \hat{p} en σ
- ✓ Stel met de formule een vergelijking op
- ✓ Los de vergelijking op met je GR.

Steekproef en betrouwbaarheidsinterval

Het 95%-betrouwbaarheidsinterval geeft aan dat bij het steeds nemen van een nieuwe aselece steekproef uit dezelfde populatie 95% van de daarbij opgestelde intervallen de populatieproportie bevat.

Het 95%-betrouwbaarheidsinterval is:

$$\checkmark [\hat{p} - 2\sigma, \hat{p} + 2\sigma]$$

Hierin is \hat{p} de steekproefproportie en σ de standaardafwijking.

Het 68%-betrouwbaarheidsinterval is:

$$\checkmark [\hat{p} - \sigma, \hat{p} + \sigma]$$

Voorbeeld

Twee dagen voor de verkiezingen voor de Tweede Kamer wordt aan 1200 stemgerechtigde Nederlanders gevraagd of ze van plan zijn te gaan stemmen. Van deze 1200 zeggen 756 inderdaad te gaan stemmen

- ✓ Bereken het 95%-betrouwbaarheidsinterval van de stemgerechtigde Nederlanders die van plan zijn te gaan stemmen.

Uitgewerkt

$$\hat{p} = \frac{756}{1200} \approx 0,63 \text{ en}$$

$$\sigma = \sqrt{\frac{0,64 \cdot 0,37}{1200}} \approx 0,013937\dots$$

Het 95%-betrouwbaarheidsinterval is [0,602; 0,658]

Voorbeeld

Bij een onderzoek vindt men een 95%-betrouwbaarheidsinterval van $[0,555; 0,605]$.

- ✓ Bereken de steekproefomvang

Uitwerking

Het interval $[0,555; 0,605]$ staat voor $[\mu - 2\sigma; \mu + 2\sigma]$, dus tussen 0,555 en 0,605 zit 4σ verschil. Dus geldt:

- ✓ $4\sigma = 0,605 - 0,555 = 0,05$ dus
 $\sigma = 0,0125$.

μ ligt in het midden van 0,555 en 0,605, dus geldt:

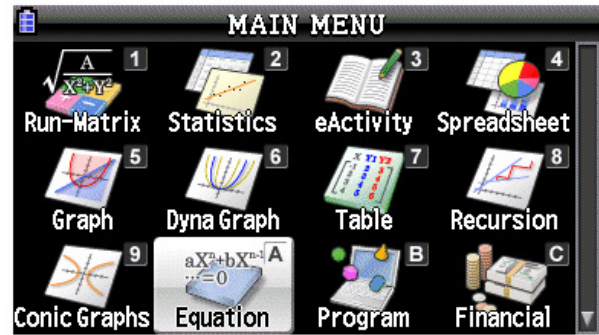
- ✓ $\mu = \frac{0,555 + 0,605}{2} = 0,58$

Vul nu de formule $\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ in:

- ✓ $\sigma = \sqrt{\frac{0,58 \cdot 0,42}{n}} = 0,0125$

Met je GR kan je dan de vergelijking oplossen.

- ✓ $n=1559$



Math Deg Norm1 ab/c Real

Equation

Select Type
F1: Simultaneous
F2: Polynomial
F3: Solver

SIMUL POLY SOLVER

Math Deg Norm1 ab/c Real

Eq: $\sqrt{\frac{.58 \times .42}{x}} = 0.0125$

$x = 1559.04$
Lft = 0.0125
Rgt = 0.0125

REPEAT

Opdracht 1 (A30)

Een fabrikant van wasmiddelen neemt een jaar (365 dagen) lang elke dag een steekproef van 117 pakken. Hieruit blijkt dat 12,0% van de pakken te weinig wasmiddel bevat.

Het percentage pakken in een dagproductie met te weinig wasmiddel is normaal verdeeld met een standaardafwijking van 3,0%.

- a. Toon aan dat dit percentage van 3,0% juist is.
- b. Bij hoeveel van de steekproeven verwacht je dat er minstens 18 pakken zijn met te weinig inhoud?
- c. Hoeveel dagen van dat jaar bevatten naar verwachting 7 of minder pakken in de steekproef te weinig wasmiddel?
- d. De bedrijfsleider beweert dat het vaak genoeg voorkomt dat in een steekproef hoogstens 2 pakken zitten met te weinig inhoud. Kan dit kloppen? Leg uit.

Denk aan de vuistregels van de normale verdeling!

Opdracht 2 (A35)

Bij de presidentsverkiezing in een land moet gekozen worden tussen de kandidaten A en B. In een peiling die voorafgaat aan de verkiezingen zeggen 935 mensen te gaan stemmen. Hiervan zeggen 487 op kandidaat A te gaan stemmen, de anderen kiezen voor kandidaat B.

- ✓ Onderzoek of je met een betrouwbaarheid van 95% kunt zeggen dat kandidaat A de verkiezingen zal gaan winnen.

Opdracht 3 (A39)

Bij een onderzoek onder D66-stemmers blijkt dat 55% het koningshuis wil behouden. Het bijbehorende betrouwbaarheidsinterval was $[0,49; 0,61]$.

- ✓ Hoeveel D66-stemmers zijn ondervraagd als het een 95%-betrouwbaarheidsinterval betreft?

Opdracht 1

a. $n=117$ en $\hat{p} = p = 0,120$ geeft:

$$\sigma = \sqrt{\frac{0,120 \cdot 0,880}{117}} \approx 0,0300\dots$$

Het percentage van 3,0% is juist.



b. Zie boven. $\frac{18}{117} \approx 0,1538\dots$. Er zit $13,5\% + 2,5\% = 16\%$ boven 0,15, Dat is dan op $0,16 \cdot 365 = 58$ dagen.

c. Zie boven. $\frac{7}{117} \approx 0,0598\dots$. Er zit 2,5% onder 0,06. Dat is dan op $0,025 \cdot 365 = 9$ dagen.

d. Zie boven. $\frac{2}{117} \approx 0,0170\dots$. Dat is zelfs kleiner dan 0,03 dus erg waarschijnlijk is het niet. De bedrijfsleider kletst maar wat...:-)

Opdracht 2

De steekproefproportie van de mensen die zeggen op A te gaan stemmen is

$$\hat{p} = \frac{487}{935} = 0,5208\dots$$

De bijbehorende standaarddeviatie is:

$$\sigma = \sqrt{\frac{0,5208\dots \cdot 0,4791\dots}{935}} = 0,0163\dots$$

Het 95%-betrouwbaarheidsinterval wordt dan:

$$[\hat{p} - 2\sigma; \hat{p} + 2\sigma]$$
$$[0,488\dots; 0,553\dots]$$

Om de verkiezingen te winnen moet kandidaat A 50% van de stemmen krijgen. Dus $p > 0,5$, maar omdat 0,488 kleiner is dan 0,5 kun je dat nu niet met 95%-zekerheid zeggen.



Opdracht 3

Bij een 95%-betrouwbaarheidsinterval weet je dat $4\sigma = 0,61 - 0,49 = 0,12$. Dus $\sigma = 0,03$.

$$\hat{p} = \frac{0,49 + 0,61}{2} = 0,55$$

Vul de formule $\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ in:

$$\sqrt{\frac{0,55 \cdot 0,45}{n}} = 0,03$$

Deze vergelijking kan je oplossen met je GR...

Er zijn 275 D66-stemmers ondervraagd.

A screenshot of a calculator screen in Math mode. It shows the equation $\text{Eq: } \sqrt{\frac{.55 \times .45}{x}} = 0.03$ and the solution $x=275$. Below the equation, it shows $\text{Lft}=0.03$ and $\text{Rgt}=0.03$. At the bottom, there is a **REPEAT** button.

Zie **F3: Solver**

4. groepen en kenmerken

Groepen vergelijken

Je kunt groepen vergelijken op basis van de kerngegevens:

- ✓ centrummaten: gemiddelde, mediaan en modus
- ✓ spreiding: standaardafwijking en kwartielafstand
- ✓ soort verdeling: symmetrisch, links- of rechtsscheef, met meerdere toppen, e.d.



Statistische samenhang en causaliteit

Bij een statistische samenhang hoeft er nog geen sprake te zijn van causaliteit (oorzaak en gevolg).

Er is alleen sprake van een oorzakelijk verband als:

- ✓ de veroorzakende variabele (de onafhankelijke variabele) in de tijd vooraf gaat aan de andere variabele (de afhankelijke variabele)
- ✓ er geen andere variabelen in het spel zijn die het verband veroorzaken

Conclusies trekken uit statistisch onderzoek

De meest voorkomende fouten:

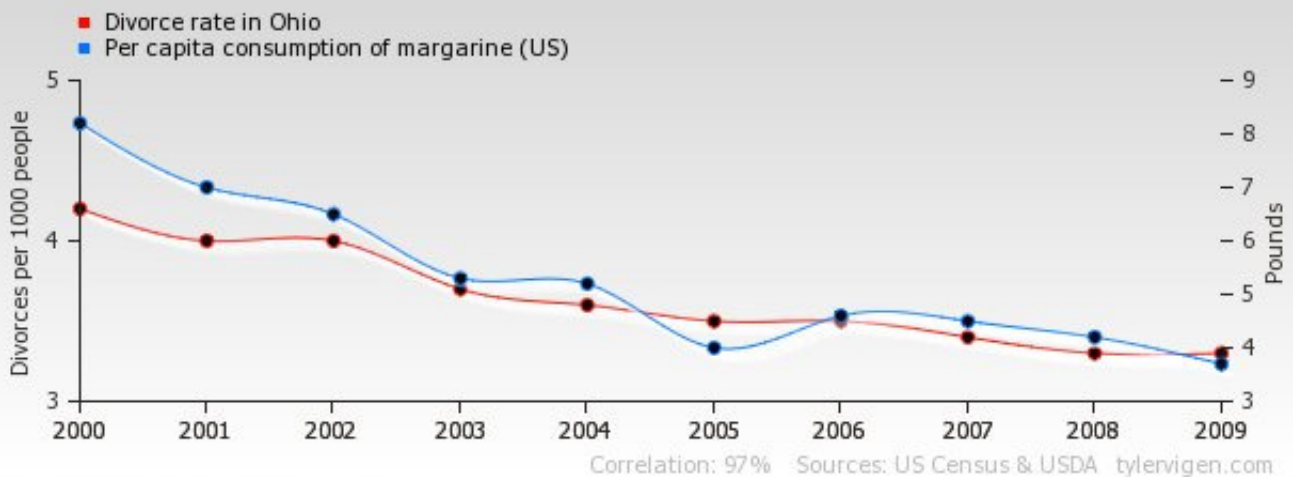
- ✓ conclusies op basis van een niet representatieve steekproef
- ✓ voor de conclusie is slechts een gedeelte van de onderzoeksresultaten gebruikt
- ✓ er wordt ten onrechte uitgegaan van causaliteit

Voorbeeld

Een klassiek voorbeeld van een correlatie zonder causaliteit is dat in gebieden waar meer ooievaars voorkomen meer kinderen worden geboren.

Een leuke website met allerlei onzincorrelatie kan je vinden op:

- ✓ **Spurious Correlations**



Opdracht 1

De minister beweert dat de werkloosheid met 3% is afgenomen. De vakbonden reageren zeer verontwaardigd, want het aantal werklozen is volgens hen toegenomen met 160.000 mensen. Toch hebben ze allebei gelijk.

- ✓ Maar kan dat eigenlijk wel? Leg uit.

Opdracht 3

Professor van der Putte heeft ontdekt dat er een grote samenhang bestaat het aantal ingezette brandweerlieden en de schade die een brand veroorzaakt. De professor adviseert om bij brand voorzichtig te zijn met het bellen van de brandweer.

- ✓ Wat denk je? Heeft de professor een punt?



Meer weten of gewoon geïnteresseerd in het onderwerp?

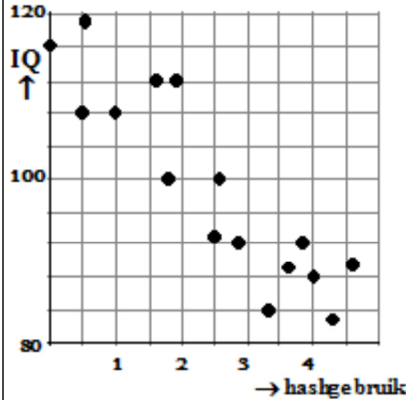
Ga 's kijken op [deze website](#)

Opdracht 2

Een onderzoeker heeft ontdekt dat als er meer ijsjes worden verkocht (variabele X) dan stijgt het aantal verdrinkingen (variabele Y).

- ✓ Geef commentaar op deze ontdekking.

Opdracht 4



In bovenstaande puntenwolk zie je de resultaten van een onderzoek naar hashgebruik en het IQ na een aantal jaren.

Welke conclusie ligt het meest voor de hand? (Geef argumenten)

- ✓ Van veel hash gebruiken gaat je IQ omlaag.
- ✓ Met een hoog IQ ga je minder hash gebruiken

Opdracht 1

Het gaat er om waar de 3% van genomen is. Als het percentage werklozen ten opzichte van de totale beroepsbevolking eerst 20% is en daarna 17% dan kun je spreken van een afname van 3%. Als de beroepsbevolking toeneemt kan het absoluut aantal werklozen toch gestegen zijn.

Voorbeeld: In een dorp wonen 100 mensen, 20 mensen zijn werkloos. Een jaar later wonen er in het dorp 200 mensen. Er zijn 34 mensen werkloos.

- ✓ Werkloosheid neemt **toe** van 20 naar 34 (absoluut).
- ✓ Werkloosheid neemt **af** van 20% naar 17% (relatief).

Opdracht 2

Het is waarschijnlijkers dat er een derde factor in het spel is. Wat dacht je van mooi weer? Dan worden er meer ijsjes verkocht (variabele X) en als meer mensen naar het strand gaan dan stijgt het aantal verdrinkingen (variabele Y) misschien ook wel...



Opdracht 3

Dat is kletspraat. Bij grotere branden is de schade groter maar er zullen ook meer brandweermannen ingezet. Een hoge samenhang maar zeker geen causaliteit.



Opdracht 4

Het lijkt erop dat veel hashgebruik inderdaad het IQ lager maakt...