

# 14 Mathematische statistiek

## Voorkennis Kansverdelingen en correlatie

### Bladzijde 50

- 1 a Het steekproefgemiddelde  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 105$

$$\text{en } \sigma_{\bar{X}} = \frac{5}{\sqrt{10}}.$$

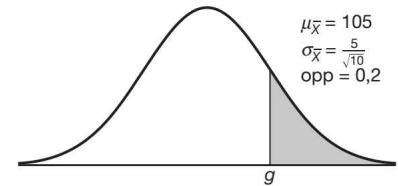
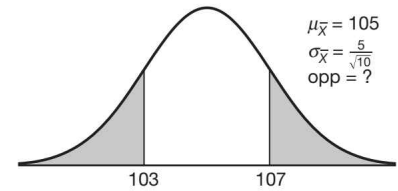
De kans is  $P(\bar{X} < 103 \vee \bar{X} > 107) = 2 \cdot P(\bar{X} < 103) =$

$$2 \cdot \text{normalcdf}\left(-10^{99}, 103, 105, \frac{5}{\sqrt{10}}\right) \approx 0,206.$$

- b opp rechts = 0,2 dus opp links = 0,8.

$$g = \text{invNorm}\left(0,8, 105, \frac{5}{\sqrt{10}}\right) = 106,33\dots$$

Het gemiddelde gewicht van de olieballen in deze zakken is 106,3 gram of meer.



### Bladzijde 51

- 2 Het steekproefgemiddelde  $\bar{L}$  is normaal verdeeld met  $\mu_{\bar{L}} = 90$  en

$$\sigma_{\bar{L}} = \frac{7}{\sqrt{12}}.$$

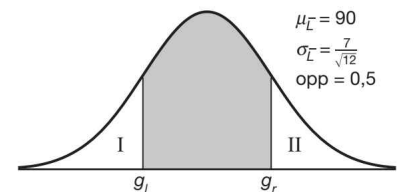
Zie de figuur.

Uit de symmetrie volgt opp I = opp II = 0,25.

$$g_l = \text{invNorm}\left(0,25, 90, \frac{7}{\sqrt{12}}\right) = 88,63\dots, \text{ dus } g_l \approx 88,6.$$

$$g_r = \text{invNorm}\left(0,75, 90, \frac{7}{\sqrt{12}}\right) = 91,36\dots, \text{ dus } g_r \approx 91,4.$$

De gemiddelde levensduur van deze pennen ligt tussen 88,6 en 91,4 meter.



- 3 a  $X$  = het aantal keer meer dan 4 ogen.

$X$  is binomiaal verdeeld met  $n = 100$  en  $p = \frac{1}{3}$ .

$$P(X \geq 40) = 1 - P(X \leq 39) = 1 - \text{binomcdf}\left(100, \frac{1}{3}, 39\right) \approx 0,097$$

- b  $P(X \geq a) < 0,06$  ofwel  $1 - P(X \leq a - 1) < 0,06$

Voer in  $y_1 = 1 - \text{binomcdf}\left(100, \frac{1}{3}, x - 1\right)$  (TI) of  $y_1 = 1 - \text{BinomialCD}\left(x - 1, 100, \frac{1}{3}\right)$  (Casio).

Uit de tabel volgt

$$x = 41 \text{ geeft } y_1 \approx 0,066$$

$$x = 42 \text{ geeft } y_1 \approx 0,043.$$

Dus  $a$  is 42, 43, 44, ..., 99 of 100.

- c  $P(X \leq b) < 0,20$

Voer in  $y_1 = \text{binomcdf}\left(100, \frac{1}{3}, x\right)$  (TI) of  $y_1 = \text{BinomialCD}\left(x, 100, \frac{1}{3}\right)$  (Casio).

Uit de tabel volgt

$$x = 28 \text{ geeft } y_1 \approx 0,152$$

$$x = 29 \text{ geeft } y_1 \approx 0,209.$$

Dus  $b$  is 0, 1, 2, ..., 27 of 28.

- d  $P(X = c) > 0,075$

Voer in  $y_1 = \text{binompdf}\left(100, \frac{1}{3}, x\right)$  (TI) of  $y_1 = \text{BinomialPD}\left(x, 100, \frac{1}{3}\right)$  (Casio).

Uit de tabel volgt

$$x = 30 \text{ geeft } y_1 \approx 0,0673$$

$$x = 31 \text{ geeft } y_1 \approx 0,0760$$

$$x = 35 \text{ geeft } y_1 \approx 0,0784$$

$$x = 36 \text{ geeft } y_1 \approx 0,0708.$$

Dus  $c$  is 31, 32, 33, 34 of 35.

**Bladzijde 54**

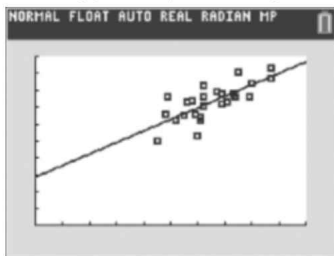
- 4** a De lengte van de langste jongen is 194 cm.  
Deze jongen is niet het zwaarst.
- b Stel  $G = aL + b$  met  $a = \frac{\Delta G}{\Delta L} = \frac{75 - 53}{190 - 160} = \frac{11}{15}$ .
- $$\left. \begin{array}{l} G = \frac{11}{15}L + b \\ \text{door } (160, 53) \end{array} \right\} \begin{array}{l} \frac{11}{15} \cdot 160 + b = 53 \\ 117\frac{1}{3} + b = 53 \\ b = -64\frac{1}{3} \end{array}$$
- Dus  $G = \frac{11}{15}L - 64\frac{1}{3}$ .
- 5** a sterke positieve correlatie  
b zwakke positieve correlatie  
c sterke negatieve correlatie
- 6** a zwakke negatieve correlatie  
b sterke positieve correlatie  
c sterke positieve correlatie  
d geen correlatie  
e zwakke positieve correlatie  
f sterke positieve correlatie  
g sterke positieve correlatie  
h geen correlatie

**14.1 Regressielijnen en correlatiecoëfficiënt****Bladzijde 55**

- 1** De lijn gaat door de punten (5, 15) en (55, 35).  
Stel  $y = ax + b$  met  $a = \frac{\Delta y}{\Delta x} = \frac{35 - 15}{55 - 5} = \frac{2}{5}$ .
- $$\left. \begin{array}{l} y = \frac{2}{5}x + b \\ \text{door } (5, 15) \end{array} \right\} \begin{array}{l} \frac{2}{5} \cdot 5 + b = 15 \\ 2 + b = 15 \\ b = 13 \end{array}$$
- Dus  $y = \frac{2}{5}x + 13$ .

**Bladzijde 56**

- 2** a Invoeren van lijst 1 = {62, 45, 58, ..., 59, 71} en lijst 2 = {83, 50, 74, ..., 66, 73} en plotten geeft het spreidingsdiagram bij onderdeel c.
- b sterke positieve correlatie
- c LinReg(ax + b) L1, L2 geeft  $\hat{Y} = 0,68X + 28,8$



- d  $X = 33$  geeft  $\hat{Y} = 0,68 \cdot 33 + 28,8 \approx 51$   
Een schatting van zijn CE-score is 51.
- e  $\hat{Y} > 55$  geeft  $0,68X + 28,8 > 55$   
 $0,68X > 26,2$   
 $X > 38,5$   
Dus vanaf een SE-score van 39.
- 3** a Voer in lijst 1 = {1, 2, 3, ..., 9, 10} en lijst 2 = {2, 3, 2, ..., 10, 12}.  
LinReg(ax + b) L1, L2 geeft  $\hat{Y} = 1,22X - 0,2$
- b  $X = 8$  geeft  $\hat{Y} = 1,22 \cdot 8 - 0,2 = 9,56$
- c Voer de residuen in bij lijst 3.  
De grootste afwijking is -1,45 bij  $X = 3$  en 1,45 bij  $X = 8$ .

d De som van de kwadraten van de residuen bereken je met  $\text{sum}(L_3^2)$ .

Je krijgt  $\text{sum}(L_3^2) \approx 8,0727$ .

e Voer in  $y_1 = 1,25x - 0,25$ .

Voer vervolgens in lijst 3 = lijst 2 -  $y_1$  (lijst 1).

Je krijgt  $\text{sum}(L_3^2) = 8,3125$ .

Dit is inderdaad groter dan in vraag d.

### Bladzijde 57

4 a Voer in lijst 1 = {6.2, 2.9, 2.9, ..., 3.4, 3.8} en lijst 2 = {36, 64, 50, ..., 47, 30}.

LinReg( $ax + b$ ) L1, L2 geeft  $\hat{Y} = -6,12X + 71,1$

b  $X = -8,0$  geeft  $\hat{Y} = -6,12 \cdot 8,0 + 71,1 \approx 120$

c  $-6,12X + 71,1 = 7$

$-6,12X = -64,1$

$X \approx 10,5$

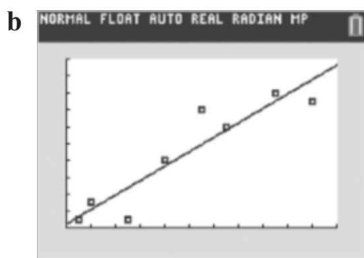
Dus vanaf een effectieve gemiddelde etmaaltemperatuur van  $10,5^\circ\text{C}$  geldt de in vraag a genoemde formule niet meer.

### Bladzijde 59

5 a Voer in lijst 1 = {5, 11, 1, ..., 17, 20} en lijst 2 = {1, 14, 1, ..., 16, 15}.

LinReg( $ax + b$ ) L1, L2 geeft  $\hat{Y} = 0,86X + 0,51$

LinReg( $ax + b$ ) L2, L1 geeft  $\hat{X} = 1,00Y + 0,85$



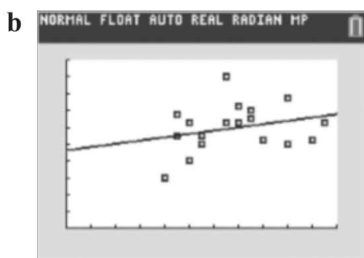
c  $Y = 10$  geeft  $\hat{X} = 1,00 \cdot 10 + 0,85 \approx 10,85$

d  $X = 10$  geeft  $\hat{Y} = 0,86 \cdot 10 + 0,51 \approx 9,11$

6 a Voer in lijst 1 = {13, 18, 14, ..., 10, 8} en lijst 2 = {36, 31, 29, ..., 16, 12}.

LinReg( $ax + b$ ) L1, L2 geeft  $\hat{Y} = 0,39X + 18,64$

LinReg( $ax + b$ ) L2, L1 geeft  $\hat{X} = 0,19Y + 8,95$



c  $Y = 18$  geeft  $\hat{X} = 0,19 \cdot 18 + 8,95 \approx 12$

d  $X = 18$  geeft  $\hat{Y} = 0,39 \cdot 18 + 18,64 \approx 26$

### Bladzijde 60

7 De verticale afstand tussen een punt en de regressielijn is  $Y - \hat{Y}$  of  $\hat{Y} - Y$ .

Dus  $(\text{verticale afstand})^2 = (Y - \hat{Y})^2 = (ax + b - y)^2$ .

Voor de regressielijnen moet  $\Sigma(\text{verticale afstand})^2$  minimaal zijn, dus  $\sum_{i=1}^n (ax_i + b - y_i)^2$  moet minimaal zijn.

8 a  $x = 3$  geeft  $\hat{Y} = 3a + b$  en  $(\text{residu})^2 = (3a + b - 5)^2$

$x = 4$  geeft  $\hat{Y} = 4a + b$  en  $(\text{residu})^2 = (4a + b - 5)^2$

$x = 5$  geeft  $\hat{Y} = 5a + b$  en  $(\text{residu})^2 = (5a + b - 5)^2$

b  $S = (a + b - 2)^2 + (2a + b - 3)^2 + (3a + b - 5)^2 + (4a + b - 5)^2 + (5a + b - 5)^2$   
 $= a^2 + b^2 + 4 + 2ab - 4a - 4b + 4a^2 + b^2 + 9 + 4ab - 12a - 6b + 9a^2 + b^2 + 25 + 6ab - 30a - 10b +$   
 $16a^2 + b^2 + 25 + 8ab - 40a - 10b + 25a^2 + b^2 + 25 + 10ab - 50a - 10b$   
 $= 55a^2 + 5b^2 + 30ab - 136a - 40b + 88$

$$\begin{aligned}
 \text{c } S \text{ is minimaal voor } b &= -\frac{30a - 40}{2 \cdot 5} = -3a + 4. \\
 \text{d } S \text{ is minimaal voor } a &= -\frac{30b - 136}{2 \cdot 55} = \frac{-30b + 136}{110}. \\
 \text{e } b = -3a + 4 \text{ invullen in } a &= \frac{-30b + 136}{110} \text{ geeft } a = \frac{-30(-3a + 4) + 136}{110} \\
 &110a = 90a - 120 + 136 \\
 &20a = 16 \\
 &a = 0,8 \\
 a = 0,8 \text{ geeft } b &= -3 \cdot 0,8 + 4 = 1,6
 \end{aligned}$$

**Bladzijde 62**

- 9 a** Stel  $\hat{Y} = aX + b$  met  $a = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$   

$$= \frac{7 \cdot 1040 - 70 \cdot 81}{7 \cdot 908 - 70^2} = 1,105\dots$$
  
 $\bar{X} = \frac{\Sigma X}{n} = \frac{70}{7} = 10$  en  $\bar{Y} = \frac{\Sigma Y}{n} = \frac{81}{7} = 11\frac{4}{7}$   

$$\left. \begin{aligned} \hat{Y} &= 1,105\dots X + b \\ (\bar{X}, \bar{Y}) &= (10, 11\frac{4}{7}) \end{aligned} \right\} \begin{aligned} 1,105\dots \cdot 10 + b &= 11\frac{4}{7} \\ b &\approx 3,83 \end{aligned}$$
  
Dus  $\hat{Y} = 1,11X + 3,83$ .
- b** Stel  $\hat{X} = aY + b$  met  $a = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{n\Sigma Y^2 - (\Sigma Y)^2}$   

$$= \frac{7 \cdot 1040 - 70 \cdot 81}{7 \cdot 1219 - 81^2} = 0,8164\dots$$
  

$$\left. \begin{aligned} \hat{X} &= 0,8164\dots Y + b \\ (\bar{Y}, \bar{X}) &= (11\frac{4}{7}, 10) \end{aligned} \right\} \begin{aligned} 0,8164\dots \cdot 11\frac{4}{7} + b &= 10 \\ b &\approx -2,45 \end{aligned}$$
  
Dus  $\hat{X} = 0,816Y - 2,45$ .
- c**  $Y = 14$  geeft  $\hat{X} = 0,816 \cdot 14 - 2,45 \approx 9,0$
- d**  $X = 14$  geeft  $\hat{Y} = 1,11 \cdot 14 + 3,83 \approx 19,4$   
Werk je met de onafgeronde regressiecoëfficiënten, dan krijg je  $\hat{Y} \approx 19,3$ .
- 10 a** Stel  $\hat{W} = aZ + b$  met  $a = \frac{n\Sigma ZW - \Sigma Z \cdot \Sigma W}{n\Sigma Z^2 - (\Sigma Z)^2}$   

$$= \frac{11 \cdot 31740,2 - 307 \cdot 1019,3}{11 \cdot 9309 - 307^2} = 4,443\dots$$
  
 $\bar{Z} = \frac{\Sigma Z}{n} = \frac{307}{11} = 27\frac{10}{11}$  en  $\bar{W} = \frac{\Sigma W}{n} = \frac{1019,3}{11} = 92\frac{73}{110}$   

$$\left. \begin{aligned} \hat{W} &= 4,443\dots Z + b \\ (\bar{Z}, \bar{W}) &= (27\frac{10}{11}, 92\frac{73}{110}) \end{aligned} \right\} \begin{aligned} 4,443\dots \cdot 27\frac{10}{11} + b &= 92\frac{73}{110} \\ b &\approx -31,4 \end{aligned}$$
  
Dus  $\hat{W} = 4,44Z - 31,4$ .
- b** Stel  $\hat{Z} = aW + b$  met  $a = \frac{n\Sigma ZW - \Sigma Z \cdot \Sigma W}{n\Sigma W^2 - (\Sigma W)^2}$   

$$= \frac{11 \cdot 31740,2 - 307 \cdot 1019,3}{11 \cdot 110836,59 - 1019,3^2} = 0,2009\dots$$
  

$$\left. \begin{aligned} \hat{Z} &= 0,2009\dots W + b \\ (\bar{W}, \bar{Z}) &= (92\frac{73}{110}, 27\frac{10}{11}) \end{aligned} \right\} \begin{aligned} 0,2009\dots \cdot 92\frac{73}{110} + b &= 27\frac{10}{11} \\ b &\approx 9,29 \end{aligned}$$
  
Dus  $\hat{Z} = 0,201W + 9,29$ .
- c**  $Z = 40$  geeft  $\hat{W} = 4,44 \cdot 40 - 31,4 \approx 146,2$   
Werk je met de onafgeronde regressiecoëfficiënten, dan krijg je  $\hat{W} \approx 146,4$ .
- d**  $W = 120$  geeft  $\hat{Z} = 0,201 \cdot 120 + 9,29 \approx 33$

**Bladzijde 63**

- 11 a Voor een punt  $(X, Y)$  in I geldt  $X > \bar{X}$  en  $Y > \bar{Y}$  dus  $X - \bar{X} > 0$  en  $Y - \bar{Y} > 0$ .  
 b  $X - \bar{X} < 0$  voor een punt in II.  
 $Y - \bar{Y} > 0$  voor een punt II.  
 c In III is  $X - \bar{X} < 0$  en  $Y - \bar{Y} < 0$  dus  $(X - \bar{X})(Y - \bar{Y}) > 0$ .  
 d Bij positieve correlatie liggen de meeste punten waarschijnlijk in I en III.

**Bladzijde 65**

- 12 a  $\sigma_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n} = \frac{0,2508}{9} \approx 0,0279$   
 b  $\frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \approx \frac{0,0279}{0,0715 \cdot 0,4841} \approx 0,805$   
 c  $\sigma_X \approx 7,15$  cm,  $\sigma_Y \approx 48,41$  cm,  $\Sigma(X - \bar{X})(Y - \bar{Y}) = 2508$  cm<sup>2</sup> en  $\sigma_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n} = \frac{2508}{9} = 278\frac{2}{3}$ .  
 d De covariatie  $\sigma_{XY}$  blijkt afhankelijk van de gebruikte eenheid.  
 e  $\frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \approx \frac{278\frac{2}{3}}{7,15 \cdot 48,41} \approx 0,805$   
 De uitkomst is hetzelfde als bij vraag b, dus  $\frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$  blijkt onafhankelijk van de gebruikte eenheid.

**Bladzijde 66**

- 13 Voer in lijst 1 = {2, 5, 7, 8, 7, 7} en lijst 2 = {4, 6, 6, 9, 7, 10}.  
 $r \approx 0,79$

**Bladzijde 67**

- 14 a 0,9  
 b 0,4  
 c -0,9
- 15 a 0,2  
 b 0,9  
 c 1
- 16 Omdat  $Y = 25 - X$  is  $\bar{Y} = 25 - \bar{X} = 25 - 18,6 = 6,4$ ,  $\sigma_Y = \sigma_X = 2,9$  en  $r = -1$ .
- 17 a 0,3  
 b 0,1  
 c 0,7

**Bladzijde 68**

- 18 a Voer in lijst 1 = {1, 2, 3, 4} en lijst 2 = {2, 2, 3, 3}.  
 $r \approx 0,89$   
 b Voer in lijst 1 = {1, 2, 3, 4} en lijst 2 = {7, 7, 8, 8}.  
 $r \approx 0,89$   
 c Voer in lijst 1 = {10, 20, 30, 40} en lijst 2 = {600, 600, 700, 700}.  
 $r \approx 0,89$
- 19 a tabel I  
 Voer in lijst 1 = {2, 3, 4, 5, 6, 7, 8} en lijst 2 = {7, 8, 6, 5, 4, 2, 3}.  
 $r \approx -0,93$   
 tabel II  
 Voer in lijst 1 = {3, 4, 6, 7, 8, 9, 5} en lijst 2 = {4, 3, 5, 9, 7, 7, 7}.  
 $r \approx 0,70$   
 tabel III  
 Voer in lijst 1 = {2, 3, 4, 5, 6, 7, 8} en lijst 2 = {8, 7, 6, 5, 4, 3, 2}.  
 $r = -1$   
 b Omdat de punten  $(X, Y)$  op de rechte lijn  $Y = 10 - X$  liggen.  
 c  $r$  verandert niet.  
 In de formule  $\sigma_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n}$  kunnen  $X$  en  $Y$  omgewisseld worden zonder dat de uitkomst verandert, en dus verandert de uitkomst van  $r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$  niet.



d  $r$  verandert niet.

In de formule  $\sigma_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n}$  verandert er niets als elke  $Y$ -waarde met 2 verminderd wordt, en ook  $\sigma_Y$  verandert hierdoor niet.

e  $r$  verandert niet.

Als elke  $X$ -waarde met tien wordt vermenigvuldigd, dan wordt  $X - \bar{X}$  tien keer zo groot en wordt

in de formule  $\sigma_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n}$  de teller tien keer zo groot, en dus wordt  $\sigma_{XY}$  tien keer zo groot.

Maar als elke  $X$ -waarde met tien wordt vermenigvuldigd, dan wordt ook  $\sigma_X$  tien keer zo groot.

In de formule  $r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$  worden teller en noemer dus beide tien keer zo groot, en dus blijft  $r$  gelijk.

### Bladzijde 69

- 20 Voer in lijst 1 = {0, 0, 0, 1, 1, 1, 2, 2, 2, 3, 3, 3}, lijst 2 = {0, 1, 2, 0, 1, 2, 0, 1, 2, 0, 1, 2} en lijst 3 = {5, 2, 0, 3, 1, 0, 0, 2, 4, 0, 0, 3}.

$r \approx 0,81$

- 21 • In het eerste spreidingsdiagram van figuur 14.24 is  $r = 0$ , maar toch is er een duidelijke samenhang zichtbaar. Deze is alleen niet lineair.
- Het tweede spreidingsdiagram van figuur 14.24 bestaat uit twee puntenwolken elk met  $r = 0$ . Bij het samenvoegen van de puntenwolken in één diagram ontstaat echter een sterk negatieve correlatie. Een mogelijke oorzaak zou kunnen zijn dat de waarnemingen uit twee verschillende deelpopulaties zijn genomen.
- Het kan gebeuren dat een van de waarnemingen er niet bij hoort. Je neemt bijvoorbeeld het maandinkomen van de aanwezigen in een klaslokaal waar de docent bij is, terwijl je eigenlijk iets wilt zeggen over het inkomen van de leerlingen. Een uitbijter zorgt voor een misleidende waarde van  $r$ . In het derde spreidingsdiagram van figuur 14.24 bijvoorbeeld is voor de vijf punten linksonder  $r = 0$ . Neem je de uitbijter mee, dan is  $r \approx 0,9$ .
- In het vierde spreidingsdiagram van figuur 14.24 is een sterk positieve correlatie zichtbaar. Je kunt bij dit diagram bijvoorbeeld denken aan een sollicitatieprocedure waarbij de kandidaten een test moeten ondergaan. Op de horizontale as staan de test scores en op de verticale as de geschiktheid voor een baan. Als je aanneemt dat de vijf kandidaten met de beste scores worden aangenomen en je zou voor deze selecte groep nogmaals de correlatie tussen de test score en de geschiktheid voor een baan onderzoeken, dan is er geen sprake meer van correlatie.
- Het is lastig de waarde van de pmcc te interpreteren, omdat bijvoorbeeld een twee keer zo grote waarde van  $r$  niet betekent dat de correlatie twee keer zo sterk is.

## 14.2 Regressievoorspellingen

### Bladzijde 71

22 a  $\sigma_X^2 = \frac{\Sigma(X - \bar{X})^2}{n} = \frac{\Sigma(X^2 - 2X\bar{X} + \bar{X}^2)}{n} = \frac{\Sigma X^2}{n} - \frac{\Sigma 2X\bar{X}}{n} + \frac{\Sigma \bar{X}^2}{n} = \frac{\Sigma X^2}{n} - 2 \frac{\Sigma X}{n} \cdot \bar{X} + \frac{n \cdot \bar{X}^2}{n} =$

$$\frac{\Sigma X^2}{n} - 2\bar{X} \cdot \bar{X} + \bar{X}^2 = \frac{\Sigma X^2}{n} - 2\bar{X}^2 + \bar{X}^2 = \frac{\Sigma X^2}{n} - \bar{X}^2$$

b  $r = \frac{\sigma_Y}{\sigma_X} = \frac{\bar{X}\bar{Y} - \bar{X} \cdot \bar{Y}}{\sigma_X^2} = \frac{\frac{\Sigma XY}{n} - \frac{\Sigma X}{n} \cdot \frac{\Sigma Y}{n}}{\frac{\Sigma X^2}{n} - \bar{X}^2} = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{n\Sigma X^2 - n^2 \cdot \bar{X}^2} = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} =$

### Bladzijde 72

- 23 a Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,4 \cdot \frac{12000}{4} = 1200$ .
- $$\left. \begin{array}{l} \hat{Y} = 1200X + b \\ (\bar{X}, \bar{Y}) = (11, 40000) \end{array} \right\} \begin{array}{l} 1200 \cdot 11 + b = 40000 \\ b = 26800 \end{array}$$
- $\hat{Y} = 1200X + 26800$
- $X = 16$  geeft  $\hat{Y} = 1200 \cdot 16 + 26800 = 46000$  dollar

b Stel  $\hat{X} = aY + b$  met  $a = r \cdot \frac{\sigma_X}{\sigma_Y} = 0,4 \cdot \frac{4}{12000} = \frac{1}{7500}$ .

$$\left. \begin{array}{l} \hat{X} = \frac{1}{7500}Y + b \\ (\bar{Y}, \bar{X}) = (40000, 11) \end{array} \right\} \frac{1}{7500} \cdot 40000 + b = 11$$

$$b = 5\frac{2}{3}$$

$$\hat{X} = \frac{1}{7500}Y + 5\frac{2}{3}$$

$Y = 35000$  geeft  $\hat{X} = \frac{1}{7500} \cdot 35000 + 5\frac{2}{3} = 10\frac{1}{3}$  jaar

24 a  $r = a_X$

$$\left. \begin{array}{l} a_X = r \cdot \frac{\sigma_X}{\sigma_Y} \end{array} \right\} \frac{\sigma_X}{\sigma_Y} = 1$$

Dus  $\sigma_X = \sigma_Y$ .

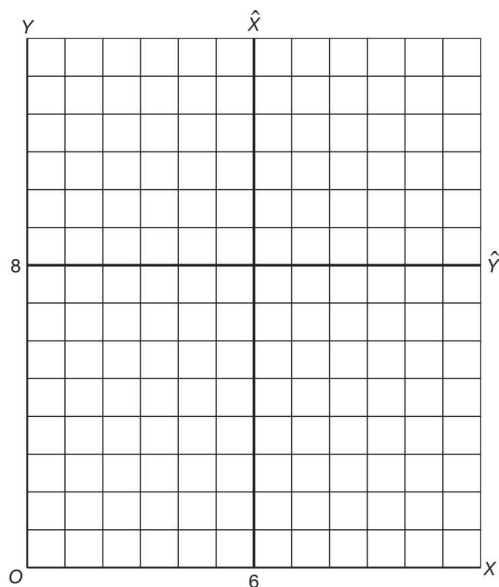
b  $r^2 = a_X \cdot a_Y = 0,8 \cdot 2,1 = 1,68 > 1$  en dat kan niet.

c  $r^2 = a_X \cdot a_Y = 3a_Y$  en  $r^2$  is maximaal 1 geeft  $a_Y$  is maximaal  $\frac{1}{3}$ .

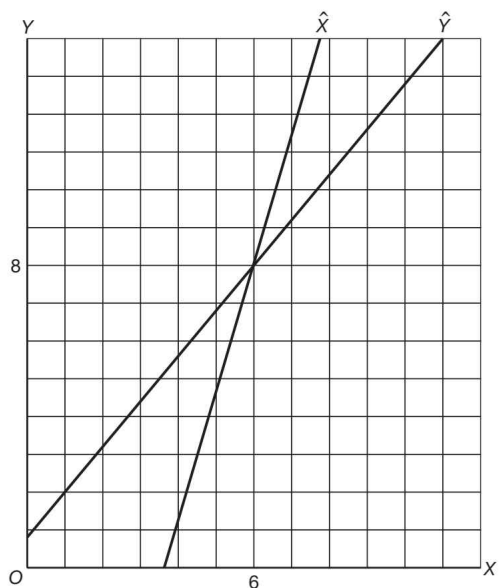
d  $a_X = -0,7$  en  $a_Y = 0,7$  geeft  $r^2 = a_X \cdot a_Y = -0,7 \cdot 0,7 = -0,49 < 0$  en dat kan niet.  
Dus dan kan  $a_Y$  niet 0,7 zijn.

e  $a_X = a_Y$  geeft  $r \cdot \frac{\sigma_X}{\sigma_Y} = r \cdot \frac{\sigma_Y}{\sigma_X}$   
Hieruit volgt  $r = 0$  of  $\sigma_X = \sigma_Y$ .

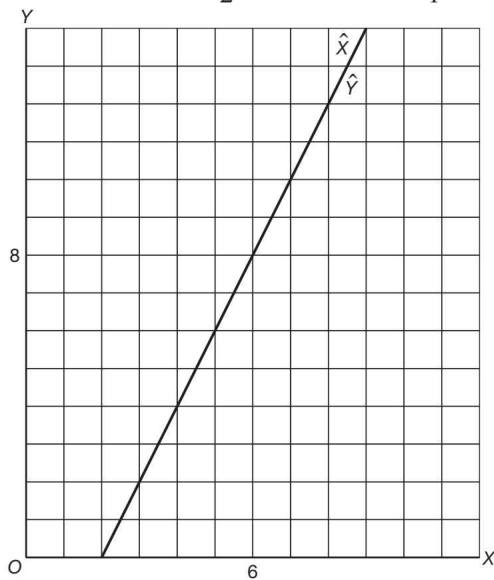
25 a  $r = 0$  geeft  $a_Y = 0$  en  $a_X = 0$ , dus een horizontale en verticale lijn door het zwaartepunt (6, 8).



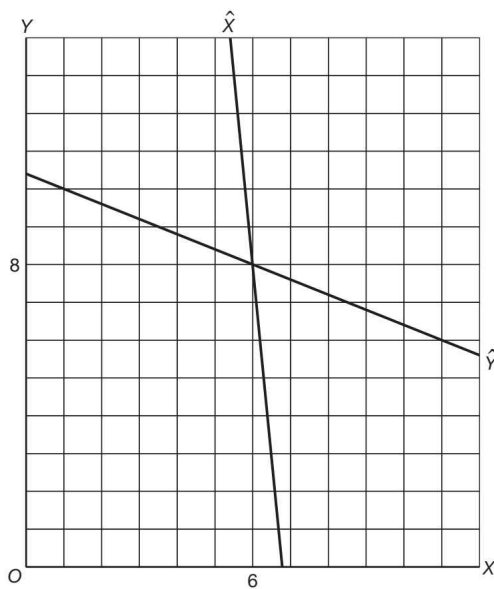
b  $r = 0,6$  geeft  $a_X = 0,6 \cdot \frac{1}{2} = 0,3$  en  $a_Y = 0,6 \cdot \frac{2}{1} = 1,2$



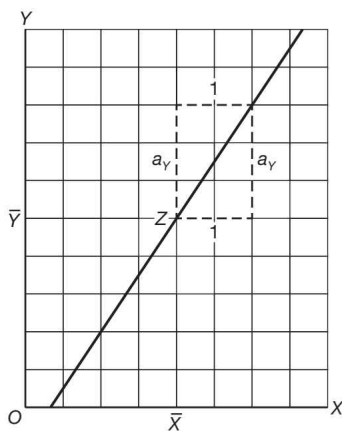
c  $r = 1$  geeft  $a_X = 1 \cdot \frac{1}{2} = 0,5$  en  $a_Y = 1 \cdot \frac{2}{1} = 2$



d  $r = -0,2$  geeft  $a_X = -0,2 \cdot \frac{1}{2} = -0,1$  en  $a_Y = -0,2 \cdot \frac{2}{1} = -0,4$



26  $r^2 = 1$  geeft  $a_X \cdot a_Y = 1$ , dus  $a_X = \frac{1}{a_Y}$ .



Vanuit  $Z(\bar{X}, \bar{Y})$  gaat de lijn  $\hat{Y} = a_Y X + b$  1 naar rechts en  $a_Y$  omhoog.

De lijn  $\hat{X} = a_X Y + b$  met  $a_X = \frac{1}{a_Y}$  gaat  $a_Y$  omhoog en 1 naar rechts vanuit  $Z(\bar{X}, \bar{Y})$ .

Dus beide lijnen vallen samen.



- 27 a**  $X$  = het IQ gemeten op twintigjarige leeftijd.  
 $Y$  = het IQ gemeten op veertigjarige leeftijd.  
 Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,80 \cdot \frac{20}{20} = 0,8$ .  

$$\left. \begin{array}{l} \hat{Y} = 0,8X + b \\ (\bar{X}, \bar{Y}) = (105, 105) \end{array} \right\} \begin{array}{l} 0,8 \cdot 105 + b = 105 \\ b = 21 \end{array}$$

$$\hat{Y} = 0,8X + 21$$
 $X = 125$  geeft  $\hat{Y} = 0,8 \cdot 125 + 21 = 121$   
**b**  $X = 95$  geeft  $\hat{Y} = 0,8 \cdot 95 + 21 = 97$

**Bladzijde 73**

- 28 a** Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,70 \cdot \frac{8}{5} = 1,12$ .  

$$\left. \begin{array}{l} \hat{Y} = 1,12X + b \\ (\bar{X}, \bar{Y}) = (65, 75) \end{array} \right\} \begin{array}{l} 1,12 \cdot 65 + b = 75 \\ b = 2,2 \end{array}$$

$$\hat{Y} = 1,12X + 2,2$$
 $X = 75$  geeft  $\hat{Y} = 1,12 \cdot 75 + 2,2 \approx 86$   
**b** Stel  $\hat{X} = aY + b$  met  $a = r \cdot \frac{\sigma_X}{\sigma_Y} = 0,70 \cdot \frac{5}{8} = 0,4375$ .  

$$\left. \begin{array}{l} \hat{X} = 0,4375Y + b \\ (\bar{Y}, \bar{X}) = (75, 65) \end{array} \right\} \begin{array}{l} 0,4375 \cdot 75 + b = 65 \\ b = 32,1875 \end{array}$$

$$\hat{X} = 0,4375Y + 32,1875$$
 $Y = 86$  geeft  $\hat{X} = 0,4375 \cdot 86 + 32,1875 \approx 70$

- 29 a** Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,85 \cdot \frac{350}{385} = \frac{17}{22}$ .  

$$\left. \begin{array}{l} \hat{Y} = \frac{17}{22}X + b \\ (\bar{X}, \bar{Y}) = (2480, 2245) \end{array} \right\} \begin{array}{l} \frac{17}{22} \cdot 2480 + b = 2245 \\ b = 328\frac{7}{11} \end{array}$$

$$\hat{Y} = \frac{17}{22}X + 328\frac{7}{11}$$
 $X = 2750$  geeft  $\hat{Y} = \frac{17}{22} \cdot 2750 + 328\frac{7}{11} \approx 2454$  gram  
**b** Stel  $\hat{X} = aY + b$  met  $a = r \cdot \frac{\sigma_X}{\sigma_Y} = 0,85 \cdot \frac{385}{350} = 0,935$ .  

$$\left. \begin{array}{l} \hat{X} = 0,935Y + b \\ (\bar{Y}, \bar{X}) = (2245, 2480) \end{array} \right\} \begin{array}{l} 0,935 \cdot 2245 + b = 2480 \\ b = 380,925 \end{array}$$

$$\hat{X} = 0,935Y + 380,925$$
 $Y = 2150$  geeft  $\hat{X} = 0,935 \cdot 2150 + 380,925 \approx 2391$  gram

**Bladzijde 74**

- 30** 
$$\left. \begin{array}{l} \hat{Y} = aX + b \\ (\bar{X}, \bar{Y}) = (6,4; 4,5) \end{array} \right\} \begin{array}{l} 6,4a + b = 4,5 \text{ ofwel } b = 4,5 - 6,4a \\ \hat{Y} = aX + b \\ (X, Y) = (8,8; 5,5) \end{array} \right\} \begin{array}{l} 8,8a + b = 5,5 \\ \text{Substitutie van } b = 4,5 - 6,4a \text{ in } 8,8a + b = 5,5 \text{ geeft } 8,8a + 4,5 - 6,4a = 5,5 \\ 2,4a = 1 \\ a = \frac{1}{2,4} \end{array}$$

$$a = r \cdot \frac{\sigma_Y}{\sigma_X} \text{ geeft } r = a \cdot \frac{\sigma_X}{\sigma_Y} = \frac{1}{2,4} \cdot \frac{0,8}{0,5} \approx 0,67$$

- 31** De voorspelling bij figuur 14.26a is betrouwbaarder omdat daar de puntenwolk dichter bij de regressielijn ligt dan in figuur 14.26b.

- 32 a**  $\text{normalcdf}(90, 110, 100, 10) \approx 0,68$ , dus 68% van de waarnemingen ligt in het interval  $\langle 90, 110 \rangle$ .  
**b**  $\mu = 33$  en  $\sigma = 9,2$  geeft  $\text{normalcdf}(23,8, 42,2, 33, 9,2) \approx 0,68$   
 $\mu = 50\,000$  en  $\sigma = 1200$  geeft  $\text{normalcdf}(48\,800, 51\,200, 50\,000, 1200) \approx 0,68$   
 $\mu = -25$  en  $\sigma = 7$  geeft  $\text{normalcdf}(-32, -18, -25, 7) \approx 0,68$   
 De bewering lijkt te kloppen.  
**c** Bij een normale verdeling met gemiddelde  $\mu$  en standaardafwijking  $\sigma$  ligt voor elke  $\mu$  en  $\sigma$  ongeveer 95% van alle waarnemingen in het interval  $\langle \mu - 2\sigma, \mu + 2\sigma \rangle$ .

**Bladzijde 76**

- 33 a** Bij  $\hat{Y} = aX + b$  zijn  $a$  en  $b$  constanten, dus geldt  $\sigma_{\hat{Y}}^2 = \sigma_{aX}^2 = (a \cdot \sigma_X)^2 = a^2 \cdot \sigma_X^2$   
 Hieruit volgt  $\sigma_{\hat{Y}} = a\sigma_X$  of  $\sigma_{\hat{Y}} = -a\sigma_X$ .  
 Omdat  $\sigma_{\hat{Y}}$  en  $\sigma_X$  positief zijn, geldt  $\sigma_{\hat{Y}} = a\sigma_X$  voor  $a > 0$  en  $\sigma_{\hat{Y}} = -a\sigma_X$  voor  $a < 0$ .  
**b**  $\left. \begin{array}{l} \sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_d^2 \\ \sigma_{\hat{Y}}^2 = a^2 \sigma_X^2 \end{array} \right\} \begin{array}{l} \sigma_Y^2 = a^2 \sigma_X^2 + \sigma_d^2 \\ \sigma_d^2 = \sigma_Y^2 - a^2 \sigma_X^2 \end{array}$   
**c** Uit vraag b en  $a = r \cdot \frac{\sigma_Y}{\sigma_X}$  volgt  $\sigma_d^2 = \sigma_Y^2 - r^2 \cdot \frac{\sigma_Y^2}{\sigma_X^2} \cdot \sigma_X^2$   

$$\sigma_d^2 = \sigma_Y^2 - r^2 \sigma_Y^2$$
  

$$\sigma_d^2 = \sigma_Y^2 (1 - r^2)$$
  

$$\sigma_d = \sigma_Y \sqrt{1 - r^2}$$
  
**34 a** Bij volkomen correlatie is  $r^2 = 1$ .  
 Dit geeft  $\sigma_d = \sigma_Y \sqrt{1 - r^2} = \sigma_Y \cdot 0 = 0$ .  
 Er is dus geen spreiding rond de modelwaarde  $\hat{Y}$ , dus alle punten liggen op de regressielijn.  
**b** Bij geen correlatie is  $r = 0$ .  
 Dit geeft  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0$ , dus  $\hat{Y} = b$ .  
 De regressielijn gaat door  $(\bar{X}, \bar{Y})$  geeft  $b = \bar{Y}$ .  
 Dus  $\hat{Y} = \bar{Y}$ .  
 Dus in het geval er geen correlatie is, is voor elke  $X$  de modelwaarde  $\hat{Y}$  gelijk aan  $\bar{Y}$ .  
 Verder is  $\sigma_d = \sigma_Y \sqrt{1 - r^2} = \sigma_Y \cdot 1 = \sigma_Y$ .  
**35 a** Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,60 \cdot \frac{10}{5} = 1,2$ .  

$$\left. \begin{array}{l} \hat{Y} = 1,2X + b \\ (\bar{X}, \bar{Y}) = (80, 65) \end{array} \right\} \begin{array}{l} 1,2 \cdot 80 + b = 65 \\ b = -31 \end{array}$$
  
 Dus  $\hat{Y} = 1,2X - 31$ .  
**b**  $\sigma_d = \sigma_Y \sqrt{1 - r^2} = 10 \sqrt{1 - 0,60^2} = 10 \sqrt{1 - 0,36} = 10 \sqrt{0,64} = 10 \cdot 0,8 = 8$   
**c**  $X = 88$  geeft  $\hat{Y} = 1,2 \cdot 88 - 31 \approx 75$   
**d** In het interval  $\langle 75 - 8, 75 + 8 \rangle = \langle 67, 83 \rangle$ .  
**e**  $X = 70$  geeft  $\hat{Y} = 1,2 \cdot 70 - 31 = 53$   
 In het interval  $\langle 53 - 2 \cdot 8, 53 + 2 \cdot 8 \rangle = \langle 37, 69 \rangle$ .

- 36 a**  $\sigma_d = \sigma_Y \sqrt{1 - r^2} = 8 \sqrt{1 - 0,90^2} \approx 3,49$   
**b** Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,90 \cdot \frac{8}{8} = 0,9$ .  

$$\left. \begin{array}{l} \hat{Y} = 0,9X + b \\ (\bar{X}, \bar{Y}) = (181, 181) \end{array} \right\} \begin{array}{l} 0,9 \cdot 181 + b = 181 \\ b = 18,1 \end{array}$$
  
 $\hat{Y} = 0,9X + 18,1$   
 $X = 184$  geeft  $\hat{Y} = 0,9 \cdot 184 + 18,1 = 183,7$  cm  
**c** In het interval  $\langle 183,7 - 2 \cdot 3,49; 183,7 + 2 \cdot 3,49 \rangle = \langle 176,7; 190,7 \rangle$ .

- 37 a Nominale schaal**  
 De schaal bestaat slechts uit een onderscheid in verschillende categorieën. De waarden van de variabelen kunnen worden aangegeven met namen zodat ze van elkaar te onderscheiden zijn.  
 Voorbeelden zijn woonplaats, beroep en geslacht.

**Ordinale schaal**

Er kan behalve onderscheid ook een volgorde worden aangegeven. De waarden op een ordinale schaal kunnen in getallen worden uitgedrukt, maar dat is niet noodzakelijk. Zo is uitmuntend beter dan goed, en goed is weer beter dan voldoende.

Voorbeelden zijn rapportcijfers, het toekennen van een aantal sterren aan een camping en het rangensysteem bij de krijgsmacht.

**Intervalschaal**

Er kan naast een volgorde ook over gelijke verschillen worden gesproken. De waarden zijn uitgedrukt in getallen waarbij het verschil tussen twee uitkomsten een eenduidige betekenis heeft.

Voorbeelden zijn temperatuur in graden Fahrenheit en de jaartelling. Op een intervalschaal is geen natuurlijk nulpunt aanwezig.

**Ratioschaal**

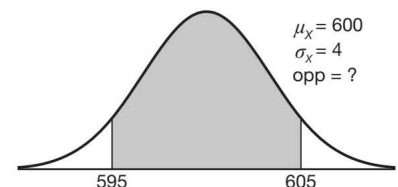
De ratioschaal heeft alle eigenschappen van bovengenoemde schaaltypen en bovendien een natuurlijk nulpunt. Dit heeft tot gevolg dat op ratioschaal verhoudingen een eenduidige betekenis hebben.

Voorbeelden zijn gewicht, leeftijd, inkomen en temperatuur in graden Kelvin.

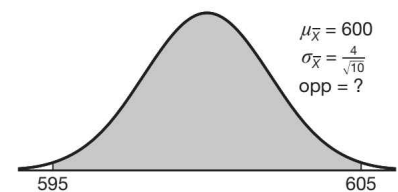
- b Bij de intervalschaal en bij de ratioschaal.  
c Bij de ordinale schaal, de intervalschaal en de ratioschaal.

**14.3 Beslissen op grond van een steekproef****Bladzijde 78**

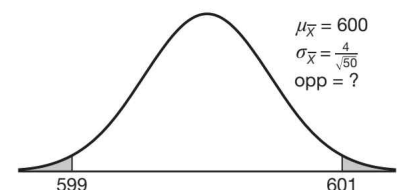
- 38 a  $X$  = de lengte in mm van een buis.  
 $X$  is normaal verdeeld met  $\mu_X = 600$  en  $\sigma_X = 4$ .  
 $P(595 < X < 605) = \text{normalcdf}(595, 605, 600, 4) \approx 0,789$



- b  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 600$  en  $\sigma_{\bar{X}} = \frac{4}{\sqrt{10}}$ .  
 $P(595 < \bar{X} < 605) = \text{normalcdf}\left(595, 605, 600, \frac{4}{\sqrt{10}}\right) \approx 0,99992$

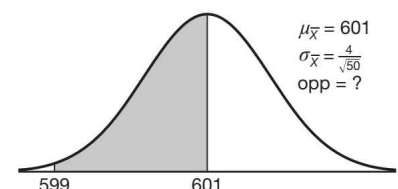
**Bladzijde 80**

- 39 a  $P(\text{ten onrechte bijstellen}) = P(\bar{X} \leq 599 \vee \bar{X} \geq 601) =$   
 $2 \cdot P(\bar{X} \leq 599) = 2 \cdot \text{normalcdf}\left(-10^9, 599, 600, \frac{4}{\sqrt{50}}\right) \approx 0,077$



- b Als Aluvia terecht tot bijstelling overgaat dan staat de machine afgesteld op een onbekend gemiddelde.  
Omdat dit gemiddelde onbekend is kun je de kans op terecht bijstellen niet berekenen.

- c  $P(\text{niet bijstellen}) = P(599 < \bar{X} < 601) =$   
 $\text{normalcdf}\left(599, 601, 601, \frac{4}{\sqrt{50}}\right) \approx 0,500$

**Bladzijde 81**

- 40 a  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 600$  en  $\sigma_{\bar{X}} = \frac{4}{\sqrt{100}} = 0,4$ .

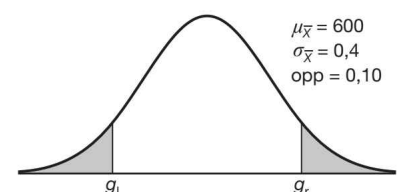
opp links = opp rechts =  $\frac{1}{2} \cdot 0,10 = 0,05$

$P(\bar{X} \leq g_l) = 0,05$ , dus

$g_l = \text{invNorm}(0,05, 600, 0,4) = 599,342... \approx 599,34$

$P(\bar{X} \leq g_r) = 1 - 0,05 = 0,95$ , dus

$g_r = \text{invNorm}(0,95, 600, 0,4) = 600,657... \approx 600,66$



- b Het steekproefgemiddelde 601,8 is groter dan  $g_r$ , dus Aluvia verworpt  $H_0$  en trekt de conclusie dat de machine moet worden bijgesteld.

**Bladzijde 82**

- 41 a  $H_0: \mu_X = 15\,000$  en  $H_1: \mu_X \neq 15\,000$

b  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 15\,000$  en  $\sigma_{\bar{X}} = \frac{600}{\sqrt{100}} = 60$ .

opp links = opp rechts =  $\frac{1}{2} \cdot 0,05 = 0,025$

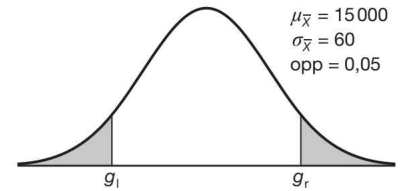
$P(\bar{X} \leq g_l) = 0,025$ , dus

$g_l = \text{invNorm}(0,025, 15\,000, 60) = 14\,882,40... \approx 14\,882,4$

$P(\bar{X} \leq g_r) = 1 - 0,025 = 0,975$ , dus

$g_r = \text{invNorm}(0,975, 15\,000, 60) = 15\,117,59... \approx 15\,117,6$

- c Het steekproefgemiddelde 14927 ligt tussen  $g_l$  en  $g_r$ , dus de fabrikant ziet geen aanleiding  $H_0$  te verwerpen en trekt de conclusie dat het nieuwe procedé de levensduur van de lampen niet beïnvloedt.

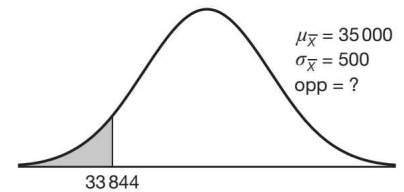


- 42 a  $H_0: \mu_X = 35\,000$  en  $H_1: \mu_X \neq 35\,000$

b  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 35\,000$  en  $\sigma_{\bar{X}} = \frac{4000}{\sqrt{64}} = 500$ .

$P(\bar{X} \leq 33\,844) = \text{normalcdf}(-10^{99}, 33\,844, 35\,000, 500) \approx 0,010$

- c  $P(\bar{X} \leq 33\,844) \approx 0,010$  en  $P(\bar{X} \leq g_l) = \frac{1}{2} \cdot 0,05 = 0,025$ , dus het steekproefgemiddelde is kleiner dan  $g_l$  en wijkt dus significant af van 35 000.

**Bladzijde 84**

- 43  $X$  = de levensduur in uren van een batterij.

$H_0: \mu_X = 2000$ ,  $H_1: \mu_X \neq 2000$  en  $\alpha = 0,05$

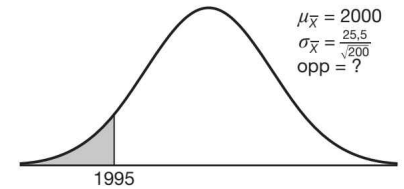
$\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 2000$  en  $\sigma_{\bar{X}} = \frac{25,5}{\sqrt{200}}$ .

De overschrijdingskans van 1995 is

$P(\bar{X} \leq 1995) = \text{normalcdf}(-10^{99}, 1995, 2000, \frac{25,5}{\sqrt{200}}) \approx 0,003$ .

$P(\bar{X} \leq 1995) \leq \frac{1}{2}\alpha$ , dus  $H_0$  wordt verworpen.

Het steekproefgemiddelde wijkt significant af van 2000.

**Bladzijde 85**

- 44 a  $H_0: \mu_X = 1,02$ ,  $H_1: \mu_X \neq 1,02$  en  $\alpha = 0,05$

$\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 1,02$  en  $\sigma_{\bar{X}} = \frac{0,04}{\sqrt{50}}$ .

De overschrijdingskans van 1,04 is

$P(\bar{X} \geq 1,04) = \text{normalcdf}(1,04, 10^{99}, 1,02, \frac{0,04}{\sqrt{50}}) \approx 0,0002$ .

$P(\bar{X} \geq 1,04) \leq \frac{1}{2}\alpha$ , dus  $H_0$  wordt verworpen.

De fabrikant zal besluiten de vulmachine opnieuw in te stellen.

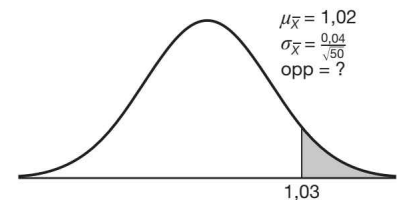
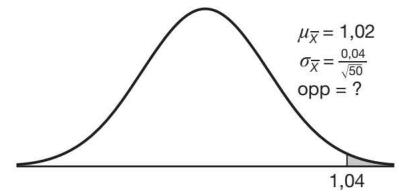
- b  $H_0: \mu_X = 1,02$ ,  $H_1: \mu_X \neq 1,02$  en  $\alpha = 0,05$

De overschrijdingskans van 1,03 is

$P(\bar{X} \geq 1,03) = \text{normalcdf}(1,03, 10^{99}, 1,02, \frac{0,04}{\sqrt{50}}) \approx 0,039$

$P(\bar{X} \geq 1,03) > \frac{1}{2}\alpha$ , dus  $H_0$  wordt niet verworpen.

De fabrikant zal de conclusie trekken dat de vulmachine niet hoeft worden bijgesteld.



- 45 a  $H_0: \mu_X = 10,2$ ,  $H_1: \mu_X \neq 10,2$  en  $\alpha = 0,10$

$\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 10,2$  en  $\sigma_{\bar{X}} = \frac{0,9}{\sqrt{40}}$ .

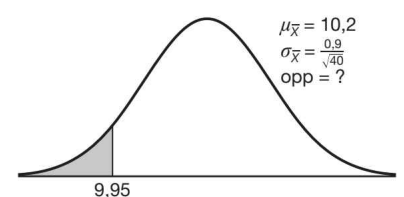
De overschrijdingskans van 9,95 is

$P(\bar{X} \leq 9,95) = \text{normalcdf}(-10^{99}, 9,95, 10,2, \frac{0,9}{\sqrt{40}}) \approx 0,039$ .

$P(\bar{X} \leq 9,95) \leq \frac{1}{2}\alpha$ , dus  $H_0$  wordt verworpen.

De afnemer trekt de conclusie dat de gemiddelde diameter significant afwijkt van 10,2 cm.

Bij  $\alpha = 0,05$  is  $P(\bar{X} \leq 9,95) \approx 0,039 > \frac{1}{2}\alpha$ . Dus  $H_0$  wordt niet verworpen. De afnemer trekt de conclusie dat de gemiddelde diameter niet significant afwijkt van 10,2 cm.





- b Het beslissingsvoorschrift heeft de vorm  
'verwerp  $H_0$  als  $\bar{X} \leq g_l$  of als  $\bar{X} \geq g_r$ '.

$\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 10,2$  en  $\sigma_{\bar{X}} = \frac{0,9}{\sqrt{100}} = 0,09$ .

opp links = opp rechts =  $\frac{1}{2} \cdot 0,10 = 0,05$

$P(\bar{X} \leq g_l) = 0,05$ , dus

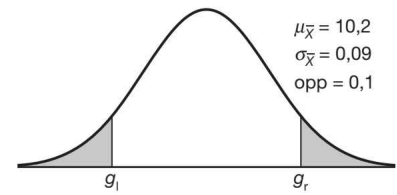
$g_l = \text{invNorm}(0,05, 10,2, 0,09) = 10,051... \approx 10,05$ .

$P(\bar{X} \leq g_r) = 1 - 0,05 = 0,95$ , dus

$g_r = \text{invNorm}(0,95, 10,2, 0,09) = 10,348... \approx 10,35$ .

Het beslissingsvoorschrift is

'verwerp  $H_0$  als  $\bar{X} \leq 10,05$  of als  $\bar{X} \geq 10,35$ '.



## 14.4 Eenzijdig en tweezijdig toetsen

### Bladzijde 87

- 46 a Omdat de bewering is dat de levensduur verlengd wordt.  
b Nee, want het steekproefgemiddelde is kleiner dan  $\mu_X = 11\,500$ .

### Bladzijde 88

- 47 a Het beslissingsvoorschrift heeft de vorm  
'verwerp  $H_0$  als  $\bar{X} \geq g$ '.

$P(\bar{X} \leq g) = 1 - 0,10 = 0,9$ , dus

$g = \text{invNorm}\left(0,9, 85, \frac{15}{\sqrt{30}}\right) \approx 88,51$ .

Het beslissingsvoorschrift is 'verwerp  $H_0$  als  $\bar{X} \geq 88,6$ '.

- b Het beslissingsvoorschrift heeft de vorm  
'verwerp  $H_0$  als  $\bar{X} \leq g$ '.

$P(\bar{X} \leq g) = 0,05$ , dus  $g = \text{invNorm}\left(0,05, 85, \frac{15}{\sqrt{50}}\right) \approx 81,51$ .

Het beslissingsvoorschrift is 'verwerp  $H_0$  als  $\bar{X} \leq 81,5$ '.

- c Het beslissingsvoorschrift heeft de vorm  
'verwerp  $H_0$  als  $\bar{X} \leq g_l$  of als  $\bar{X} \geq g_r$ '.

opp links = opp rechts =  $\frac{1}{2} \cdot 0,01 = 0,005$

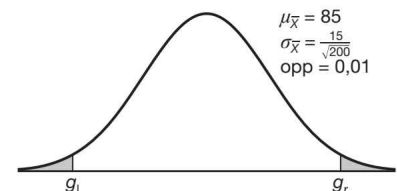
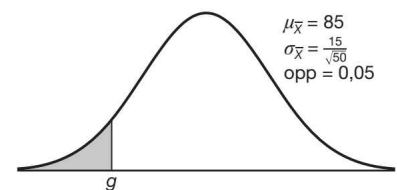
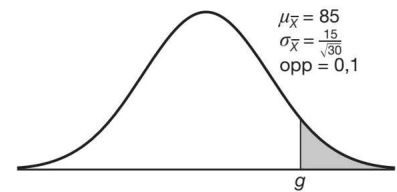
$P(\bar{X} \leq g_l) = 0,005$ , dus

$g_l = \text{invNorm}\left(0,005, 85, \frac{15}{\sqrt{200}}\right) \approx 82,27$ .

$P(\bar{X} \leq g_r) = 0,995$ , dus

$g_r = \text{invNorm}\left(0,995, 85, \frac{15}{\sqrt{200}}\right) \approx 87,73$ .

Het beslissingsvoorschrift is 'verwerp  $H_0$  als  $\bar{X} \leq 82,2$  of als  $\bar{X} \geq 87,8$ '.



- 48 a  $X$  = de afhandelingstijd in minuten van een bestelling.

$H_0: \mu_X = 12$ ,  $H_1: \mu_X < 12$  en  $\alpha = 0,05$

$\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 12$  en  $\sigma_{\bar{X}} = \frac{3}{\sqrt{25}} = 0,6$ .

$P(\bar{X} \leq g) = 0,05$ , dus  $g = \text{invNorm}(0,05, 12, 0,6) \approx 11,01$ .

Dus bij steekproefgemiddelden van 11,0 minuten of minder is er aanleiding te veronderstellen dat de afhandelingstijd verminderd is.

- b  $H_0: \mu_X = 12$ ,  $H_1: \mu_X < 12$  en  $\alpha = 0,01$

$\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 12$  en  $\sigma_{\bar{X}} = \frac{3}{\sqrt{80}}$ .

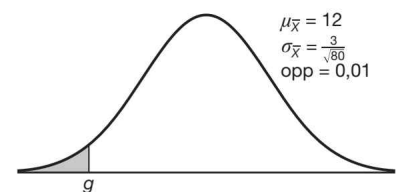
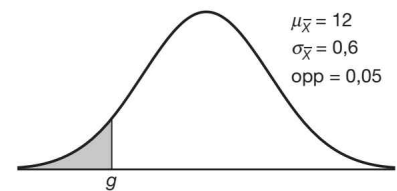
$P(\bar{X} \leq g) = 0,01$ , dus  $g = \text{invNorm}\left(0,01, 12, \frac{3}{\sqrt{80}}\right) \approx 11,21$ .

Het beslissingsvoorschrift is 'verwerp  $H_0$  als  $\bar{X} \leq 11,2$ '.

Het steekproefgemiddelde is 11 minuten en 18 seconden = 11,3 minuten.

Dit is meer dan 11,2 dus  $H_0$  wordt niet verworpen.

Er is geen aanleiding om te concluderen dat de afhandelingstijd is afgenomen.





**Bladzijde 90**

- 49  $X$  = het gewicht in gram van een pakje margarine.

$$H_0: \mu_X = 500, H_1: \mu_X > 500 \text{ en } \alpha = 0,05$$

Het steekproefgemiddelde  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 500$  en  $\sigma_{\bar{X}} = \frac{1,5}{\sqrt{100}} = 0,15$ .

De overschrijdingskans van 500,4 is  $P(\bar{X} \geq 500,4) = \text{normalcdf}(500,4, 10^{99}, 500, 0,15) \approx 0,004$ .

$P(\bar{X} \geq 500,4) \leq \alpha$ , dus verwerp  $H_0$ .

Er is reden om de productieafdeling gelijk te geven.

- 50 a  $P(\text{tablet werkt goed}) = P(3,8 < X < 4,2) = \text{normalcdf}(3,8, 4,2, 4, 0,12) \approx 0,904$

- b  $H_0: \mu_X = 4, H_1: \mu_X \neq 4$  en  $\alpha = 0,05$

Het steekproefgemiddelde  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 4$  en  $\sigma_{\bar{X}} = \frac{0,12}{\sqrt{50}}$ .

De overschrijdingskans van 3,95 is  $P(\bar{X} \leq 3,95) = \text{normalcdf}(-10^{99}, 3,95, 4, \frac{0,12}{\sqrt{50}}) \approx 0,002$ .

$P(\bar{X} \leq 3,95) \leq \frac{1}{2}\alpha$ , dus verwerp  $H_0$ .

Het gemiddelde wijkt significant af van 4 mg.

- c  $P(\text{tablet werkt niet goed}) = P(X \leq 3,8 \vee X \geq 4,2) = P(X \leq 3,8) + P(X \geq 4,2) = \text{normalcdf}(-10^{99}, 3,8, 3,95, 0,12) + \text{normalcdf}(4,2, 10^{99}, 3,95, 0,12) \approx 0,124$

Dus 12,4% van de tabletten werkt dan niet goed.

- 51 a  $X$  = de inhoud in gram van een pak koffie.

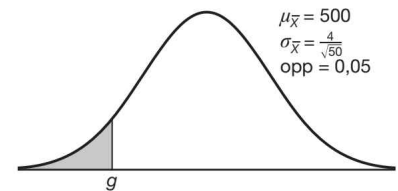
$$H_0: \mu_X = 500, H_1: \mu_X < 500 \text{ en } \alpha = 0,05$$

Het steekproefgemiddelde  $\bar{X}$  is normaal verdeeld met

$$\mu_{\bar{X}} = 500 \text{ en } \sigma_{\bar{X}} = \frac{4}{\sqrt{50}}$$

$$P(\bar{X} \leq g) = 0,05, \text{ dus } g = \text{invNorm}\left(0,05, 500, \frac{4}{\sqrt{50}}\right) \approx 499,07.$$

Dus het gemiddelde gewicht per pak moet 499,0 gram of kleiner zijn.



- b  $H_0: \mu_X = 500, H_1: \mu_X > 500$  en  $\alpha = 0,025$

Het steekproefgemiddelde  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 500$  en  $\sigma_{\bar{X}} = \frac{4}{\sqrt{25}} = 0,8$ .

De overschrijdingskans van 501,94 is  $P(\bar{X} \geq 501,94) = \text{normalcdf}(501,94, 10^{99}, 500, 0,8) \approx 0,008$ .

$P(\bar{X} \geq 501,94) \leq \alpha$ , dus verwerp  $H_0$ .

Het hoofd van de afdeling Voorraad krijgt gelijk.

- c  $H_0: \mu_X = 500, H_1: \mu_X \neq 500$  en  $\alpha = 0,05$

Het steekproefgemiddelde  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 500$  en  $\sigma_{\bar{X}} = \frac{4}{\sqrt{25}} = 0,8$ .

De overschrijdingskans van 501,48 is  $P(\bar{X} \geq 501,48) = \text{normalcdf}(501,48, 10^{99}, 500, 0,8) \approx 0,032$

$P(\bar{X} \geq 501,48) > \frac{1}{2}\alpha$ , dus verwerp  $H_0$  niet.

Het steekproefresultaat wijkt niet significant af van 500 gram.

**Bladzijde 91**

- 52 a  $X$  = de gemiddelde trekkracht in newton.

De overschrijdingskans is  $P(\bar{X} \leq 785)$ .

Als dit steekproefresultaat bij  $H_0: \mu = 800$  aanleiding geeft om  $H_0$  te verwerpen is

$$P(\bar{X} \leq 785) \leq \alpha.$$

Omdat  $P(\bar{X} \leq 785)$  kleiner wordt als  $\mu$  groter dan 800 wordt, is ook  $P(\bar{X} \leq 785) \leq \alpha$

bij  $H_0: \mu = a$  met  $a > 800$ .

Dus wordt  $H_0$  ook verworpen bij  $\mu > 800$ .

- b Bij  $\mu = 800$  is de overschrijdingskans  $P(\bar{X} \leq 785)$  groter dan bij een  $\mu$ -waarde groter dan 800.

Bij  $\mu = 800$  wordt  $H_0$  dus minder snel verworpen dan bij een  $\mu$ -waarde groter dan 800, dus dit is gunstiger voor de fabrikant.

- 53  $X$  = het aantal uur dat een Nederlander per week tv kijkt.

$$H_0: \mu_X = 23,4, H_1: \mu_X < 23,4 \text{ en } \alpha = 0,025$$

Het steekproefgemiddelde  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 22,6$  en  $\sigma_{\bar{X}} = \frac{2,4}{\sqrt{30}}$ .

De overschrijdingskans van 22,6 is  $P(\bar{X} \leq 22,6) = \text{normalcdf}(-10^{99}, 22,6, 23,4, \frac{2,4}{\sqrt{30}}) \approx 0,034$

$P(\bar{X} \leq 22,6) > \alpha$ , dus verwerp  $H_0$  niet.

Er is geen aanleiding om het eens te zijn met de bewering van de medewerker van De Ster.

## 14.5 Binomiale toetsen

### Bladzijde 93

- 54 a Je let alleen op de gebeurtenissen succes (de persoon vindt de frisdrank van Mol de lekkerste) en mislukking. Verder is bij elk kansexperiment de kans op succes hetzelfde. Als Mol gelijk heeft, dan is  $p = 0,40$ .
- b Nee, want  $X = 48$  is meer dan het verwachte aantal van 40 dat hoort bij  $p = 0,40$ .
- c De concurrent krijgt dan gelijk, want  $X = 28$  ligt ver onder het verwachte aantal van 40 als  $p = 0,40$ .

### Bladzijde 94

- 55  $X$  = het aantal mensen dat later met een allergie te kampen kreeg.
- $X$  is binomiaal verdeeld met  $n = 474$  en  $p = 0,3$ .
- $$H_0: p = 0,3, H_1: p < 0,3 \text{ en } \alpha = 0,025$$
- De overschrijdingskans van 112 is  $P(X \leq 112) = \text{binomcdf}(474, 0,3, 112) \approx 0,001$ .
- $P(X \leq 112) \leq \alpha$ , dus verwerp  $H_0$ .
- Er is aanleiding de onderzoekers gelijk te geven.

### Bladzijde 95

- 56  $X$  = het aantal mannen dat aan een vorm van kleurenblindheid lijdt.
- $X$  is binomiaal verdeeld met  $n = 200$  en  $p = 0,08$ .
- $$H_0: p = 0,08, H_1: p > 0,08 \text{ en } \alpha = 0,05$$
- De overschrijdingskans van 22 is
- $$P(X \geq 22) = 1 - P(X \leq 21) = 1 - \text{binomcdf}(200, 0,08, 21) \approx 0,080.$$
- $P(X \geq 22) > \alpha$ , dus verwerp  $H_0$  niet.
- Er is geen aanleiding om bioloog Vosseberg gelijk te geven.

- 57  $X$  = het aantal keer zes ogen bij het gooien met een dobbelsteen.
- $X$  is binomiaal verdeeld met  $n = 80$  en  $p = \frac{1}{6}$ .
- $$H_0: p = \frac{1}{6}, H_1: p < \frac{1}{6} \text{ en } \alpha = 0,05$$
- De overschrijdingskans van 8 is  $P(X \leq 8) = \text{binomcdf}(80, \frac{1}{6}, 8) \approx 0,067$ .
- $P(X \leq 8) > \alpha$ , dus verwerp  $H_0$  niet.
- Er is niet voldoende aanleiding om het met Mirjam eens te zijn.

- 58  $X$  = het aantal keer op rood bij het draaien van de schijf.
- $X$  is binomiaal verdeeld met  $n = 160$  en  $p = \frac{1}{4}$ .
- $$H_0: p = \frac{1}{4}, H_1: p > \frac{1}{4} \text{ en } \alpha = 0,01$$
- De overschrijdingskans van 52 is
- $$P(X \geq 52) = 1 - P(X \leq 51) = 1 - \text{binomcdf}(160, \frac{1}{4}, 51) \approx 0,020.$$
- $P(X \geq 52) > \alpha$ , dus verwerp  $H_0$  niet.
- Er is niet voldoende reden om Simon gelijk te geven.

**Bladzijde 96**

- 59**  $X$  = het aantal patiënten waarbij het middel een positieve uitwerking had.  
 $X$  is binomiaal verdeeld met  $n = 500$  en  $p = 0,80$ .  
 $H_0: p = 0,80, H_1: p < 0,80$  en  $\alpha = 0,05$   
 Verwerp  $H_0$  als  $X \leq g$ .  
 Zoek het grootste gehele getal  $g$  zo, dat  $P(X \leq g) \leq 0,05$  ofwel  $\text{binomcdf}(500, 0,80, g) \leq 0,05$ .  
 Voer in  $y_1 = \text{binomcdf}(500, 0,80, x)$  (TI) of  $y_1 = \text{BinomialCD}(x, 500, 0,80)$  (Casio).  
 Uit de tabel volgt  
 $x = 384$  geeft  $y_1 \approx 0,0434$   
 $x = 385$  geeft  $y_1 \approx 0,0543$ .  
 Verwerp  $H_0$  als  $X \leq 384$ .  
 Het kleinste aantal patiënten waarbij de bewering van de fabrikant niet wordt verworpen is 385.

- 60 a**  $X$  = het aantal sets dat gewonnen wordt door een speler die de set met serveren begint.  
 $X$  is binomiaal verdeeld met  $n = 500$  en  $p = 0,55$ .  
 $H_0: p = 0,55, H_1: p < 0,55$  en  $\alpha = 0,05$   
 Verwerp  $H_0$  als  $X \leq g$ .  
 Zoek het grootste gehele getal  $g$  zo, dat  $P(X \leq g) \leq 0,05$  ofwel  $\text{binomcdf}(500, 0,55, g) \leq 0,05$ .  
 Voer in  $y_1 = \text{binomcdf}(500, 0,55, x)$  (TI) of  $y_1 = \text{BinomialCD}(x, 500, 0,55)$  (Casio).  
 Uit de tabel volgt  
 $x = 256$  geeft  $y_1 \approx 0,04837$   
 $x = 257$  geeft  $y_1 \approx 0,05805$ .  
 Verwerp  $H_0$  als  $X \leq 256$ .  
 Het aantal sets waarbij de speler die begint met serveren de set wint is hoogstens 256.
- b**  $X = 242 \leq 256$ , dus verwerp  $H_0$ .  
 Jacco krijgt gelijk. Het door de tennisliefhebber genoemde percentage is te optimistisch.
- c**  $Y$  = het aantal games dat wordt gewonnen door een speler die met nieuwe ballen heeft mogen serveren.  
 $Y$  is binomiaal verdeeld met  $n = 300$  en  $p = 0,815$ .  
 $H_0: p = 0,815, H_1: p > 0,815$  en  $\alpha = 0,025$   
 Verwerp  $H_0$  als  $Y \geq g$ .  
 Zoek het kleinste gehele getal  $g$  zo, dat  $P(Y \geq g) \leq 0,025$  ofwel  $1 - P(Y \leq g - 1) \leq 0,025$   
 ofwel  $1 - \text{binomcdf}(300, 0,815, g - 1) \leq 0,025$ .  
 Voer in  $y_1 = 1 - \text{binomcdf}(300, 0,815, x - 1)$  (TI) of  $y_1 = 1 - \text{BinomialCD}(x - 1, 300, 0,815)$  (Casio).  
 Uit de tabel volgt  
 $x = 257$  geeft  $y_1 \approx 0,03415$   
 $x = 258$  geeft  $y_1 \approx 0,02375$ .  
 Verwerp  $H_0$  als  $Y \geq 258$ .  
 De serveerder moet minstens 258 games winnen opdat Jacco gelijk krijgt.

**Bladzijde 98**

- 61 a** Een onzuiver geldstuk kan te veel maar ook te weinig keer kop geven.
- b**  $X$  = het aantal keer kop bij het gooien met een geldstuk.  
 $X$  is binomiaal verdeeld met  $n = 100$  en  $p = 0,5$ .  
 $H_0: p = 0,5, H_1: p \neq 0,5$  en  $\alpha = 0,05$   
 De overschrijdingskans van 59 is  
 $P(X \geq 59) = 1 - P(X \leq 58) = 1 - \text{binomcdf}(100, 0,5, 58) \approx 0,044$ .  
 $P(X \geq 59) > \frac{1}{2}\alpha$ , dus verwerp  $H_0$  niet.  
 Conclusie: het geldstuk is zuiver.
- 62**  $X$  = het aantal keer zes ogen bij het gooien met een dobbelsteen.  
 $X$  is binomiaal verdeeld met  $n = 150$  en  $p = \frac{1}{6}$ .  
 $H_0: p = \frac{1}{6}, H_1: p \neq \frac{1}{6}$  en  $\alpha = 0,05$   
 De overschrijdingskans van 12 is  $P(X \leq 12) = \text{binomcdf}(150, \frac{1}{6}, 12) \approx 0,002$ .  
 $P(X \leq 12) \leq \frac{1}{2}\alpha$ , dus verwerp  $H_0$ .  
 Conclusie: de dobbelsteen is niet zuiver.



**Bladzijde 99**

- 63 a**  $X$  = het aantal keer dat het balletje op sector 1 blijft liggen.  
 $X$  is binomiaal verdeeld met  $n = 500$  en  $p = 0,2$ .  
 $H_0: p = 0,2, H_1: p \neq 0,2$  en  $\alpha = 0,01$   
 De overschrijdingskans van 115 is  
 $P(X \geq 115) = 1 - P(X \leq 114) = 1 - \text{binomcdf}(500, 0,2, 114) \approx 0,054$ .  
 $P(X \geq 115) > \frac{1}{2}\alpha$ , dus verwerp  $H_0$  niet.  
 Er is geen aanleiding te vermoeden dat de roulette niet zuiver is.
- b**  $X$  = het aantal keer dat het balletje op de sectoren 4 en 5 blijft liggen.  
 $X$  is binomiaal verdeeld met  $n = 600$  en  $p = 0,4$ .  
 $H_0: p = 0,4, H_1: p \neq 0,4$  en  $\alpha = 0,01$   
 Verwerp  $H_0$  als  $X \leq g_l$  of als  $X \geq g_r$ .  
 Zoek het grootste gehele getal  $g_l$  zo, dat  $P(X \leq g_l) \leq \frac{1}{2}\alpha = 0,005$  ofwel  $\text{binomcdf}(600, 0,4, g_l) \leq 0,005$ .  
 En zoek het kleinste gehele getal  $g_r$  zo, dat  $P(X \geq g_r) \leq \frac{1}{2}\alpha = 0,005$  ofwel  $1 - P(X \leq g_r - 1) \leq 0,005$   
 ofwel  $1 - \text{binomcdf}(600, 0,4, g_r - 1) \leq 0,005$ .  
 Voer in  $y_1 = \text{binomcdf}(600, 0,4, x)$  (TI) of  $y_1 = \text{BinomialCD}(x, 600, 0,4)$  (Casio).  
 Uit de tabel volgt  
 $x = 208$  geeft  $y_1 \approx 0,0041$   
 $x = 209$  geeft  $y_1 \approx 0,0052$ .  
 Verwerp  $H_0$  als  $X \leq 208$ .  
 Voer in  $y_1 = 1 - \text{binomcdf}(600, 0,4, x - 1)$  (TI) of  $y_1 = 1 - \text{BinomialCD}(x - 1, 600, 0,4)$  (Casio).  
 Uit de tabel volgt  
 $x = 271$  geeft  $y_1 \approx 0,0057$   
 $x = 272$  geeft  $y_1 \approx 0,0045$ .  
 Verwerp  $H_0$  als  $X \geq 272$ .  
 De zuiverheid van de roulette zal niet in twijfel worden getrokken bij de aantallen 209 tot en met 271.
- c**  $X$  = het aantal keer dat het balletje op de sectoren met een even getal blijft liggen.  
 $X$  is binomiaal verdeeld met  $n = 300$  en  $p = 0,4$ .  
 $H_0: p = 0,4, H_1: p < 0,4$  en  $\alpha = 0,025$   
 De overschrijdingskans van 110 is  $P(X \leq 110) = \text{binomcdf}(300, 0,4, 110) \approx 0,131$ .  
 $P(X \leq 110) > \alpha$ , dus verwerp  $H_0$  niet.  
 Er is niet voldoende reden om het met Erik eens te zijn.
- 64**  $X$  = het aantal spaarlampen met een levensduur van meer dan 8000 uur.  
 $X$  is binomiaal verdeeld met  $n = 30$  en  $p = 0,8$ .  
 $H_0: p = 0,8, H_1: p < 0,8$  en  $\alpha = 0,10$   
 Het steekproefresultaat is 21 op een totaal van 30.  
 De overschrijdingskans van 21 is  $P(X \leq 21) = \text{binomcdf}(30, 0,8, 21) \approx 0,129$ .  
 $P(X \leq 21) > \alpha$ , dus verwerp  $H_0$  niet.  
 De steekproef geeft voldoende aanleiding de fabrikant in het gelijk te stellen.
- 65 a**  $X$  = het aantal baby's dat behoort tot de categorie 'zwaar'.  
 $X$  is binomiaal verdeeld met  $n = 80$  en  $p = \text{normalcdf}(4000, 10^9, 3250, 425) = 0,03880...$   
 $P(X \geq 5) = 1 - P(X \leq 4) = 1 - \text{binomcdf}(80, 0,03880..., 4) \approx 0,199$
- b**  $H_0: p = 0,0388, H_1: p > 0,0388$  en  $\alpha = 0,01$   
 De overschrijdingskans van 8 is  
 $P(X \geq 8) = 1 - P(X \leq 7) = 1 - \text{binomcdf}(58, 0,0388..., 7) \approx 0,002$ .  
 $P(X \geq 8) \leq \alpha$ , dus verwerp  $H_0$ .  
 Conclusie: je kunt de medewerkers van het consultatiebureau gelijk geven.



**Bladzijde 100**

- 66 a** Je bekijkt voor elk nat jaar eenmalig of het in de periode 1901-1950 valt, of in de periode 1951-2000. Dus je hebt te maken met het kansmodel van trekken zonder terugleggen. Stel als volgt de vaas samen: 100 knikkers waarvan 50 rood (de jaren 1901 t/m 1950) en 50 wit (de jaren 1961 t/m 2000). Als de 10 natste jaren willekeurig zijn verdeeld over deze 100 jaren, dan is de kans dat 9 van de 10

natste jaren in de periode 1951-2000 vallen, en 1 natste jaar in de periode 1901-1950 dus  $\frac{\binom{50}{1} \cdot \binom{50}{9}}{\binom{100}{10}}$ .

De overschrijdingskans van 9 is  $\frac{\binom{50}{1} \cdot \binom{50}{9}}{\binom{100}{10}} + \frac{\binom{50}{10}}{\binom{100}{10}} \approx 0,008$ .

$0,008 \leq 0,01$ , dus er is sprake van een significante afwijking.

De stelling dat toevallige variaties in neerslag ervoor zorgen dat negen van tien natste jaren in de periode 1901-2000 in de jaren na 1950 vallen, wordt verworpen.

Je kunt dus concluderen dat extreem natte jaren na 1950 vaker voorkomen.

- b**  $X$  = het aantal van de tien natste jaren in de periode 1901-1980.

$X$  is binomiaal verdeeld met  $n = 10$  en  $p = 0,8$ .

$H_0: p = 0,8$ ,  $H_1: p \neq 0,8$  en  $\alpha = ?$

Uit de tabel volgt  $X = 5$ , en omdat  $np = 10 \cdot 0,8 = 8$  bereken je  $P(X \leq 5) = \text{binomcdf}(10, 0,8, 5) = 0,0327...$

De toets is tweezijdig. Dit geeft  $\alpha = 2 \cdot 0,0327... \approx 0,066$ .

- c** Je hebt te maken met het kansmodel van trekken zonder terugleggen, dus het gebruik van de binomiale verdeling is niet toegestaan.

- d**  $H_0$ : in de periode 1961-2000 komen extreem natte jaren gemiddeld even vaak voor als in de periode 1901-1960,

$H_1$ : in de periode 1961-2000 komen extreem natte jaren gemiddeld vaker voor dan in de periode 1901-1960,  $\alpha = 0,01$

$Y$  = het aantal van de tien natste jaren in de periode 1961-2000.

Uit de tabel volgt  $Y = 8$ .

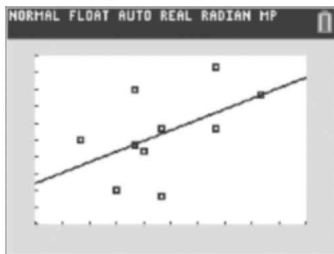
De overschrijdingskans van 8 is  $P(Y \geq 8) = \frac{\binom{40}{8} \cdot \binom{60}{2} + \binom{40}{9} \cdot \binom{60}{1} + \binom{40}{10}}{\binom{100}{10}} \approx 0,00886$ .

$0,00886 \leq \alpha$ , dus verwerp  $H_0$ .

De onderzoeker krijgt gelijk.

**Diagnostische toets****Bladzijde 102**

- 1 a** Voer in lijst 1 = {164, 155, 161, ..., 175, 162} en lijst 2 = {50, 60, 69, ..., 68, 58}.



- b** LinReg( $ax + b$ ) L1, L2 geeft  $\hat{Y} = 0,629X - 42,0$

LinReg( $ax + b$ ) L2, L1 geeft  $\hat{X} = 0,409Y + 139$

- c**  $\hat{Y} > 75$

$$0,629X - 42,0 > 75$$

$$0,629X > 117$$

$$X > 186,0$$

Dus vanaf 186 cm.

- d  $\hat{X} > 170$   
 $0,409Y + 139 > 170$   
 $0,409Y > 31$   
 $Y > 75,8$   
 Dus vanaf 76 kg.

2 a Stel  $\hat{Y} = aX + b$  met  $a = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$   

$$= \frac{5 \cdot 63\,360,8 - 5523 \cdot 60,7}{5 \cdot 6\,567\,097 - 5523^2} = -0,007908...$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{5523}{5} = 1104,6 \text{ en } \bar{Y} = \frac{\Sigma Y}{n} = \frac{60,7}{5} = 12,14$$

$$\left. \begin{array}{l} \hat{Y} = -0,007908...X + b \\ (\bar{X}, \bar{Y}) = (1104,6; 12,14) \end{array} \right\} \begin{array}{l} -0,007908... \cdot 1104,6 + b = 12,14 \\ b \approx 20,9 \end{array}$$

Dus  $\hat{Y} = -0,00791X + 20,9$ .

b Stel  $\hat{X} = aY + b$  met  $a = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{n\Sigma Y^2 - (\Sigma Y)^2}$   

$$= \frac{5 \cdot 63\,360,8 - 5523 \cdot 60,7}{5 \cdot 766,67 - 60,7^2} = -123,8...$$

$$\left. \begin{array}{l} \hat{X} = -123,8...Y + b \\ (\bar{Y}, \bar{X}) = (12,14; 1104,6) \end{array} \right\} \begin{array}{l} -123,8... \cdot 12,14 + b = 1104,6 \\ b \approx 2609 \end{array}$$

Dus  $\hat{X} = -124Y + 2609$ .

c  $X = 1000$  geeft  $\hat{Y} = -0,00791 \cdot 1000 + 20,9 \approx 13 \text{ km}$

- 3 Voer in lijst 1 = {7, 9.5, 11, 9, 9.5, 8, 8, 10}, lijst 2 = {169, 171, 178, 160, 174, 192, 168, 182} en lijst 3 = {56, 65, 78, 56, 53, 60, 57, 82}.
- Bij  $X$  en  $Y$  is  $r \approx 0,142$ .
- Bij  $X$  en  $Z$  is  $r \approx 0,695$ .
- Bij  $Y$  en  $Z$  is  $r \approx 0,429$ .
- Dus  $X$  en  $Z$  vertonen de sterkste correlatie.

4 a Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,8 \cdot \frac{1}{2} = 0,4$ .

$$\left. \begin{array}{l} \hat{Y} = 0,4X + b \\ (\bar{X}, \bar{Y}) = (10, 12) \end{array} \right\} \begin{array}{l} 0,4 \cdot 10 + b = 12 \\ b = 8 \end{array}$$

$$\hat{Y} = 0,4X + 8$$

$X = 13$  geeft  $\hat{Y} = 0,4 \cdot 13 + 8 = 13,2$

b  $a_X = r \cdot \frac{\sigma_X}{\sigma_Y}$   
 $0,6 = r \cdot \frac{2}{1}$   
 $r = 0,3$

5 a Stel  $\hat{Y} = aX + b$  met  $a = r \cdot \frac{\sigma_Y}{\sigma_X} = 0,6 \cdot \frac{3}{10} = 0,18$ .

$$\left. \begin{array}{l} \hat{Y} = 0,18X + b \\ (\bar{X}, \bar{Y}) = (100, 20) \end{array} \right\} \begin{array}{l} 0,18 \cdot 100 + b = 20 \\ b = 2 \end{array}$$

Dus  $\hat{Y} = 0,18X + 2$ .

b  $\sigma_d = \sigma_Y \sqrt{1 - r^2} = 3 \sqrt{1 - 0,6^2} = 3 \sqrt{1 - 0,36} = 3 \sqrt{0,64} = 3 \cdot 0,8 = 2,4$

c  $X = 85$  geeft  $\hat{Y} = 0,18 \cdot 85 + 2 = 17,3$   
 Het gevraagde interval is  $\langle 17,3 - 2,4; 17,3 + 2,4 \rangle = \langle 14,9; 19,7 \rangle$ .

**Bladzijde 103**

- 6 a**  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 2200$  kg en  $\sigma_{\bar{X}} = \frac{250}{\sqrt{20}}$  kg.  
 $P(\text{ten onrechte het gemiddelde als afwijkend beschouwen}) =$

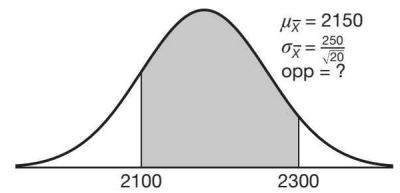
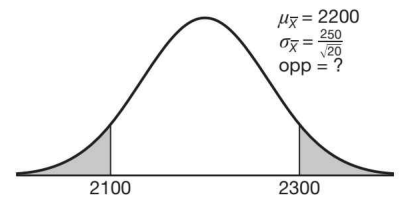
$$P(\bar{X} \leq 2100) + P(\bar{X} \geq 2300) = 2 \cdot P(\bar{X} \leq 2100) =$$

$$2 \cdot \text{normalcdf}\left(-10^{99}, 2100, 2200, \frac{250}{\sqrt{20}}\right) \approx 0,074$$

- b**  $\bar{X}$  is normaal verdeeld met  $\mu_{\bar{X}} = 2150$  kg en  $\sigma_{\bar{X}} = \frac{250}{\sqrt{20}}$  kg.

$$P(\text{conclusie gemiddelde is niet afwijkend}) =$$

$$\text{normalcdf}\left(2100, 2300, 2150, \frac{250}{\sqrt{20}}\right) \approx 0,811$$



- 7**  $X$  = het gewicht in kg van een zak aardappelen.

$$H_0: \mu_X = 5, H_1: \mu_X < 5 \text{ en } \alpha = 0,025$$

$$\text{Het steekproefgemiddelde } \bar{X} \text{ is normaal verdeeld met } \mu_{\bar{X}} = 5 \text{ en } \sigma_{\bar{X}} = \frac{0,4}{\sqrt{50}}.$$

$$\text{De overschrijdskans van 4,76 is } P(\bar{X} \leq 4,76) = \text{normalcdf}\left(-10^{99}, 4,76, 5, \frac{0,4}{\sqrt{50}}\right) \approx 0,00001.$$

$$P(\bar{X} \leq 4,76) \leq \alpha, \text{ dus verwerp } H_0.$$

Je kunt zeggen dat de consumentenorganisatie gelijk krijgt.

- 8 a**  $X$  = het aantal keer dat een van de getallen 6 tot en met 11 wordt aangewezen.

$$X \text{ is binomiaal verdeeld met } n = 100 \text{ en } p = 0,3.$$

$$H_0: p = 0,3, H_1: p \neq 0,3 \text{ en } \alpha = 0,05$$

$$\text{Verwerp } H_0 \text{ als } X \leq g_l \text{ of als } X \geq g_r.$$

$$\text{Zoek het grootste gehele getal } g_l \text{ zo, dat } P(X \leq g_l) \leq \alpha = 0,025 \text{ ofwel } \text{binomcdf}(100, 0,3, g_l) \leq 0,025.$$

$$\text{En zoek het kleinste gehele getal } g_r \text{ zo, dat } P(X \geq g_r) \leq 0,025 \text{ ofwel } 1 - P(X \leq g_r - 1) \leq 0,025$$

$$\text{ofwel } 1 - \text{binomcdf}(100, 0,3, g_r - 1) \leq 0,025.$$

$$\text{Voer in } y_1 = \text{binomcdf}(100, 0,3, x) \text{ (TI) of } y_1 = \text{BinomialCD}(x, 100, 0,3) \text{ (Casio).}$$

Uit de tabel volgt

$$x = 20 \text{ geeft } y_1 \approx 0,0165$$

$$x = 21 \text{ geeft } y_1 \approx 0,0288.$$

$$\text{Verwerp } H_0 \text{ als } X \leq 20.$$

$$\text{Voer in } y_1 = 1 - \text{binomcdf}(100, 0,3, x - 1) \text{ (TI) of } y_1 = 1 - \text{BinomialCD}(x - 1, 100, 0,3) \text{ (Casio).}$$

Uit de tabel volgt

$$x = 39 \text{ geeft } y_1 \approx 0,0340$$

$$x = 40 \text{ geeft } y_1 \approx 0,0210.$$

$$\text{Verwerp } H_0 \text{ als } X \geq 40.$$

Dus bij de aantallen 21 tot en met 39 zal de zuiverheid van het rad niet in twijfel worden getrokken.

- b**  $H_0: p = 0,3, H_1: p \neq 0,3 \text{ en } \alpha = 0,01$

$$\text{De overschrijdskans van 70 is } P(X \geq 70) = 1 - P(X \leq 69) = 1 - \text{binomcdf}(200, 0,3, 69) \approx 0,073.$$

$$P(X \geq 70) > \frac{1}{2}\alpha, \text{ dus verwerp } H_0 \text{ niet.}$$

Er is geen aanleiding te vermoeden dat het rad niet zuiver is.

- 9 a**  $P(\text{een pak koffie weegt minder dan 1000 gram}) = \text{normalcdf}(-10^{99}, 1000, 1006, 5) = 0,1150... \approx 0,115.$

- b**  $X$  = het aantal pakken dat minder dan 1000 gram weegt.

$$X \text{ is binomiaal verdeeld met } n = 250 \text{ en } p = 0,1150...$$

$$H_0: p = 0,115, H_1: p < 0,115 \text{ en } \alpha = 0,01$$

$$\text{De overschrijdskans van 18 is } P(X \leq 18) = \text{binomcdf}(250, 0,1150..., 18) \approx 0,017.$$

$$P(X \leq 18) > \alpha, \text{ dus verwerp } H_0 \text{ niet.}$$

Er is niet aangetoond dat het gemiddelde is toegenomen.