

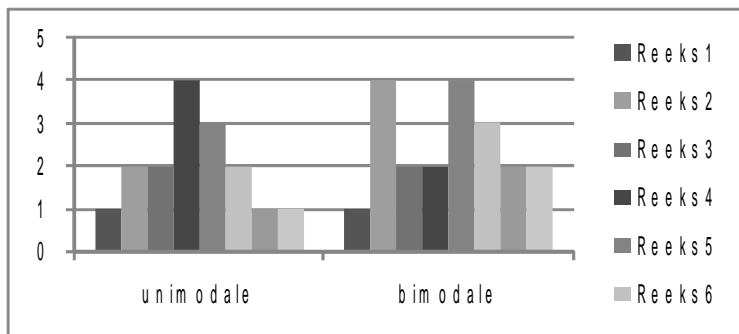
Bijlage bij Statistics: The arts of learning from data (Agresti & Finlay)

Hoofdstuk 2

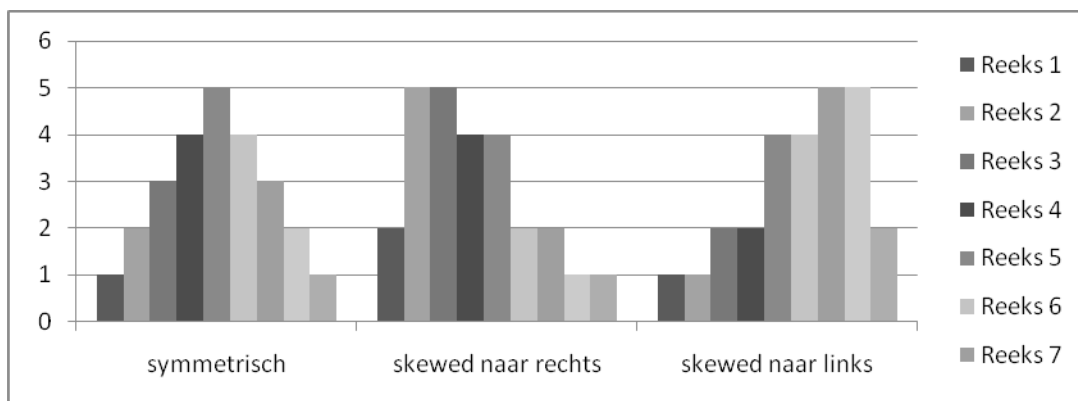
Figuur 1

Stem	Leaves
6	127999
7	335669
8	00144467

Figuur 2



Figuur 3



Figuur 4

Data zonder uitschieter: 1 2 2 2 2 2 3 4 5 5 6 7 7 7 8 9 10
Gemiddelde: $88/17 = 4.82$
Mediaan = 5
Modus: 2 (er zijn vijf 2)

Data met uitschieter: 1 2 2 2 2 2 3 4 5 5 6 7 7 7 8 9 23
Gemiddelde: $101/17 = 5.94$
Mediaan = 5
Modus: 2

Formule gemiddelde:

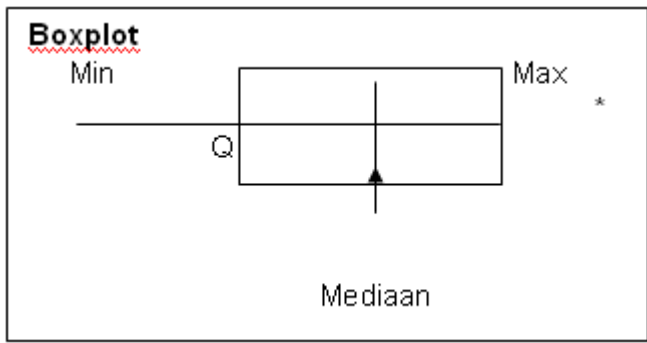
$$\bar{x} = \frac{\sum x}{n}$$

Formule 1: $(x - \bar{x})$

Formule 2:

Formule variantie :	Formule standaarddeviatie:
$s^2 = \frac{\sum(x - \bar{x})^2}{n-1}$	$S = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} = \sqrt{s^2}$

Figuur 5



Figuur 6

Formule Z-score:

$$Z\text{-score} = \frac{X - \mu}{\sigma}$$

Hoofdstuk 3

Figuur 7

Contingency table			
Willen trouwen			
Sekse	Ja	Nee	Totaal
Man	130 (0.38)	210 (0.62)	340
Vrouw	303 (0.76)	95 (0.24)	398
Totaal	433	305	738

Van de mannen wil 0.38 (130 : 340 = 0.38) wel trouwen en 0.62 niet. Bij de vrouwen is dit 0.76 wel en 0.24 niet.

Er is dus een associatie tussen sekse en het wel of niet willen trouwen.

Formule 3:

Formule correlatie, deze hoeft voorlopig niet gebruikt te worden. Al is het wel handig om deze al een keer gezien te hebben.

$$r = \frac{1}{n-1} \sum \left(\frac{x-\bar{x}}{Sx} \right) \left(\frac{y-\bar{y}}{Sy} \right)$$

Formule 4:

Formule regressie-vergelijking:

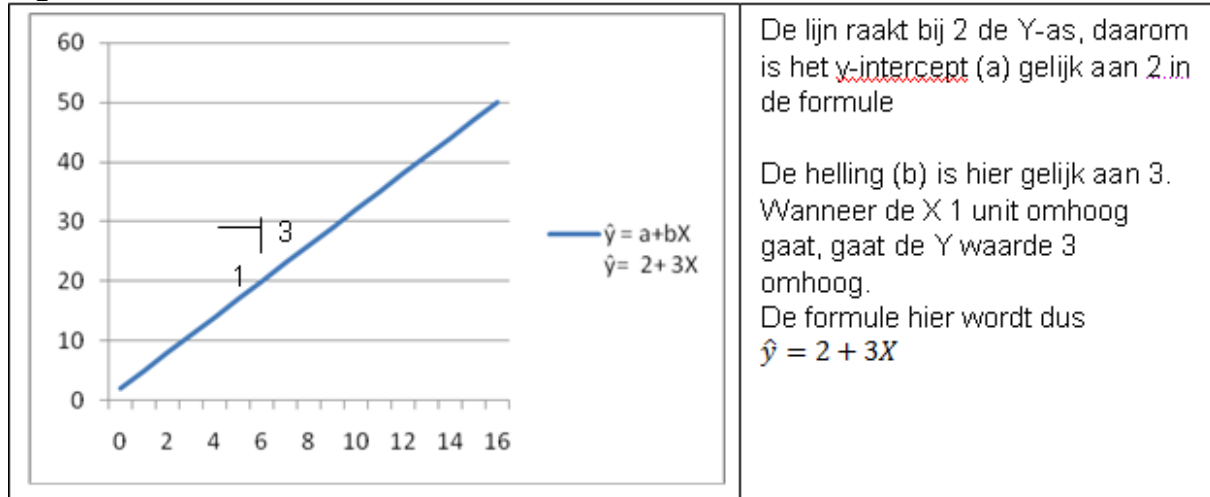
$$\hat{y} = a + bX$$

\hat{y} = staat voor de voorspelde waarde van y

Aanduiding 1: a

Aanduiding 2: b

Figuur 8



Figuur 9

Formule residual sum of squares:
$$\sum(\text{residuen})^2 = \sum(y - \hat{y})^2$$

Aanduiding 3: (\bar{X}, \bar{Y})

Formule 5

Y-intercept = $\mathbf{a} = \bar{Y} - b(\bar{X})$	r = de correlatie
<u>helling</u> = $\mathbf{b} = r \left(\frac{S_y}{S_x} \right)$	<u>S_y</u> = standaarddeviatie van y
	<u>S_x</u> = standaarddeviatie van x

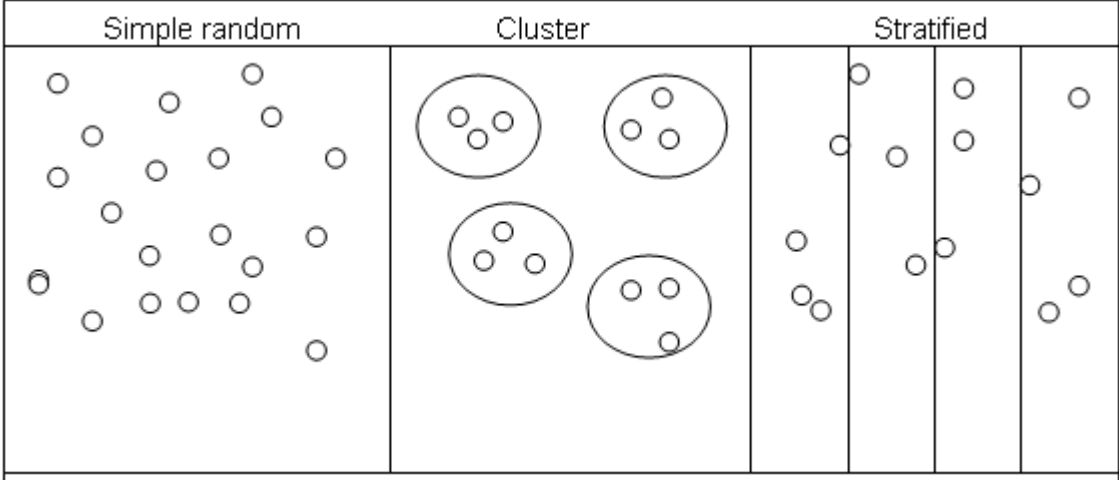
Hoofdstuk 4

Formule 6:

Voorlopige formule margin of error:

$$\frac{1}{\sqrt{n}} \times 100\%$$

Figuur 10



Hoofdstuk 5

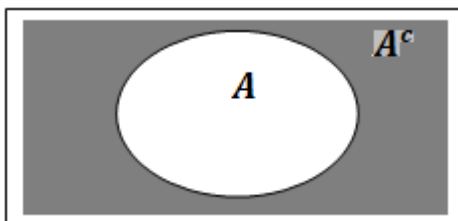
Figuur 11

$$P(A) = \frac{\text{Aantal uitkomsten in event } A}{\text{Aantal mogelijke uitkomsten in sample space}}$$

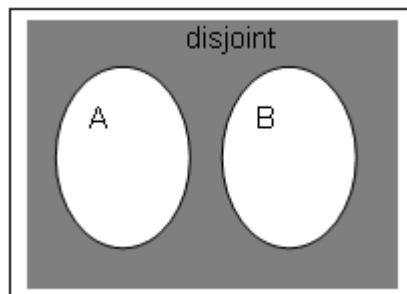
Formule 7: $P(A) = \frac{4}{6}$

Formule 8: $P(A^c) = 1 - P(A)$

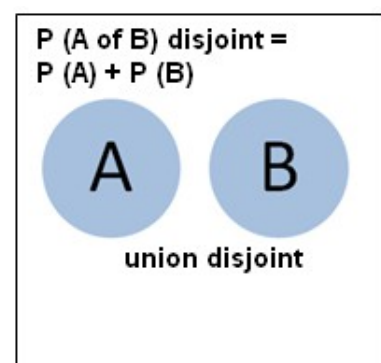
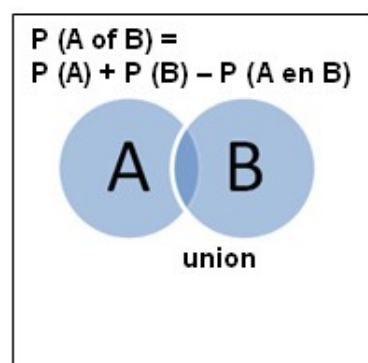
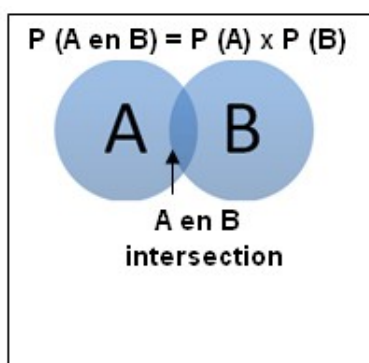
Figuur 12



Figuur 13



Figuur 14



Formule 9

$$P(A|B) = \frac{P(A \text{ en } B)}{P(B)} \quad \text{of} \quad P(B|A) = \frac{P(A \text{ en } B)}{P(A)}$$

B. $P(B|A)$ = de kans op B

Hoofdstuk 6

Formule 10:

$$\begin{array}{l} \text{Formule Z-score:} \\ \text{Z-score} = \frac{X - \mu}{\sigma} \end{array}$$

Formule 11:

$$P(x) = p^x \cdot (1-p)^{n-x} \cdot \binom{n}{x}$$

Formule 12: $p^x = \frac{1^2}{6}$

Formule 13: $(1-p)^{n-x} = (1 - \frac{1}{6})^{3-2}$

Formule 14: $\binom{n}{x} = \binom{3}{2}$

Formule 15:

$$\begin{array}{l} P(2) = \frac{1^2}{6} \cdot (1 - \frac{1}{6})^{3-2} \cdot \binom{3}{2} = 0.069 \\ P(2) = \frac{1^2}{6} \cdot \frac{5^1}{6} \cdot \binom{3}{2} = 0.069 \end{array}$$

Formule 16:

$$\text{Gemiddelde} = n \cdot p \qquad \text{Standaarddeviatie} = \sqrt{n \cdot p (1-p)}$$

Formule 17:

$$\text{Gemiddelde} = p \qquad \text{Standaarddeviatie} = \sqrt{\frac{p(1-p)}{n}}$$

Aanduiding 4: \bar{x}

Formule 18:

$$\text{Standaard error van } \bar{x} = \frac{\sigma}{\sqrt{n}}$$

Formule 19: $\sqrt{\frac{p(1-p)}{n}}$

Voor kwantitatieve variabelen heeft de sampling distributie van de sample gemiddelde \bar{x} , een gemiddelde dat gelijk staat aan het populatiegemiddelde μ en een standaarderror van $\frac{\sigma}{\sqrt{n}}$. De σ staat hierbij voor de standaarddeviatie van de populatie.

Hoofdstuk 7

Formule 20:

Margin of error:
Sample proportie: $1,96 \cdot SE$

Formule 21:

Formule betrouwbaarheidsinterval populatieproportie:
 $\hat{p} \pm (1,96 (se))$ = geschatte waarde \pm margin of error
 $Se = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Formule 22: het aantal successen ($n \cdot \hat{p}$) ≥ 15 en het aantal niet successen ($n(1-\hat{p})$) ≥ 15 .

Formule 23: $\hat{p} \pm 2,58 \cdot (se)$

Aanduiding 5: \bar{X}

Formule 24:

Formule betrouwbaarheidsinterval populatiegemiddelde:
 $\bar{x} \pm (t_{0,025} (se))$ = Geschatte waarde \pm margin of error
 $Se = \frac{s}{\sqrt{n}}$

Formule 25:

Formule sample size berekening populatieproportie:
 $n = \frac{\hat{p}(1-\hat{p})z^2}{m^2}$

Aanduiding 6: \hat{p}

Formule 26:

Formule sample size berekening populatiegemiddelde:
 $n = \frac{4s^2}{m^2}$

Hoofdstuk 8

Formule 27:

$$z = \frac{\hat{p} - p_0}{se_0} \quad se_0 = \sqrt{\frac{p_0(1-p_0)}{n}}$$

Formule 28:

$$t = \frac{\bar{x} - \mu_0}{se} \quad se = \frac{s}{\sqrt{n}}$$

Figuur 15

	<i>Beslissing</i>	
<i>werkelijkheid</i>	H₀ niet verwerpen	H₀ verwerpen
H₀ is waar	Juiste beslissing	Type 1 fout
H₀ is niet waar	Type 2 fout	Juiste beslissing (power)

Hoofdstuk 9

Formule 29:

Betrouwbaarheidsinterval proportie van twee groepen:

$$P1 - P2 \pm Z (se) \quad se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Formule 30:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se0} \quad se0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

\hat{p} in $se0$ = de poolestimate = totaal aantal successen van beide samples / totaal aantal observaties van beide samples

Formule 31: = \hat{p}_1 / \hat{p}_2

Formule 32:

Betrouwbaarheidsinterval gemiddelde van twee groepen:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{.025} (se) \quad se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Formule 33:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se} \quad se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Formule 34: \bar{x}_1 / \bar{x}_2

Aanduiding 7: \bar{x}_d

Hoofdstuk 10

Formule 35:

$$\chi^2 = \sum \frac{(\text{geobserveerde data} - \text{verwachte data})^2}{\text{verwachte data}}$$

Figuur 16

Geobserveerde waarde is de waarde die uit het onderzoek is verkregen.

<i>Geobserveerde waarde</i>	mavo	havo	vwo	Totaal
Laag inkomen	45	38	20	103
Gemiddeld inkomen	40	45	43	128
Hoog inkomen	18	26	50	94
Totaal	103	109	113	325

De **verwachte waarde** moet zelf berekend worden. Dit doet men door de verwachte uitkomst per cel te berekenen: (rij totaal x kolom totaal) delen door totale sample size. De verwachte waarde voor laag inkomen en mavo is $(103 \times 103) / 325 = 32,6$

<i>verwachte waarde</i>	mavo	havo	vwo	Totaal
Laag inkomen	32,6	34,5	35,8	103
Gemiddeld inkomen	40,6	42,9	44,5	128
Hoog inkomen	29,7	31,5	32,7	94
Totaal	103	109	113	325

Formule 36:

$$\frac{(\text{geobserveerde waarde} - \text{verwachte waarde})^2}{se}$$

De formule van se is complex en wordt met behulp van software berekend.

Hoofdstuk 11

Formule 37:

$$\begin{aligned} \text{Regressielijn: } & \hat{y} = a + bx \\ \text{y-intercept: } & a = \bar{y} - b\bar{x} \\ \text{helling/slope : } & b = r \left(\frac{S_y}{S_x} \right) \end{aligned}$$

Formule 38:

$$\text{Formule: sum of squared residuals: } \sum(\text{residu})^2 = \sum(y - \hat{y})^2$$

Formule 39:

$$r = b \left(\frac{S_y}{S_x} \right), \text{ wanneer } S_y = S_x, \text{ dan } r = b$$

Aanduiding 8: \hat{y}

Aanduiding 9: \bar{y}

Formule 40:

$$\mu_y = \alpha + \beta x$$

Formule 41:

$$t = \frac{b - 0}{se} \quad \text{Se en b = verkregen via software}$$

Formule 42:

$$b \pm t_{.025} (se)$$

Hoofdstuk 12

Formule 43:

$$\mu_y = \alpha + \beta_1 x_1 + \beta_2 x_2$$

Formule 44:

$$t = \frac{b_1 - 0}{se} \quad \text{zowel } b_1 \text{ als se wordt door software geleverd.}$$

Formule 45:

$$F = \frac{\text{mean square for regression}}{\text{mean square error}} = \frac{\text{regressie SS} / \# \text{ variabele (=df1)}}{\text{residu SS} / n - \# \text{ parameters (=df2)}}$$

Formule 46:

$$\begin{array}{ll} \text{Vrouw: } \mu_y = \alpha + \beta(1) & \mu_y = 1,2 + 0,58 \cdot (1) = 1,78 \\ \text{Man: } \mu_y = \alpha + \beta(0) & \mu_y = 1,2 + 0,58 \cdot (0) = 1,2 \end{array}$$

Hoofdstuk 13

Formule 47:

$$F = \frac{\text{Between groups variance}}{\text{within groups variance}} = \frac{\text{variantie tussen de groepen}}{\text{variantie in de groep}}$$

Aanduiding 10: (\bar{y}_1 ; met sample grootte n_1)

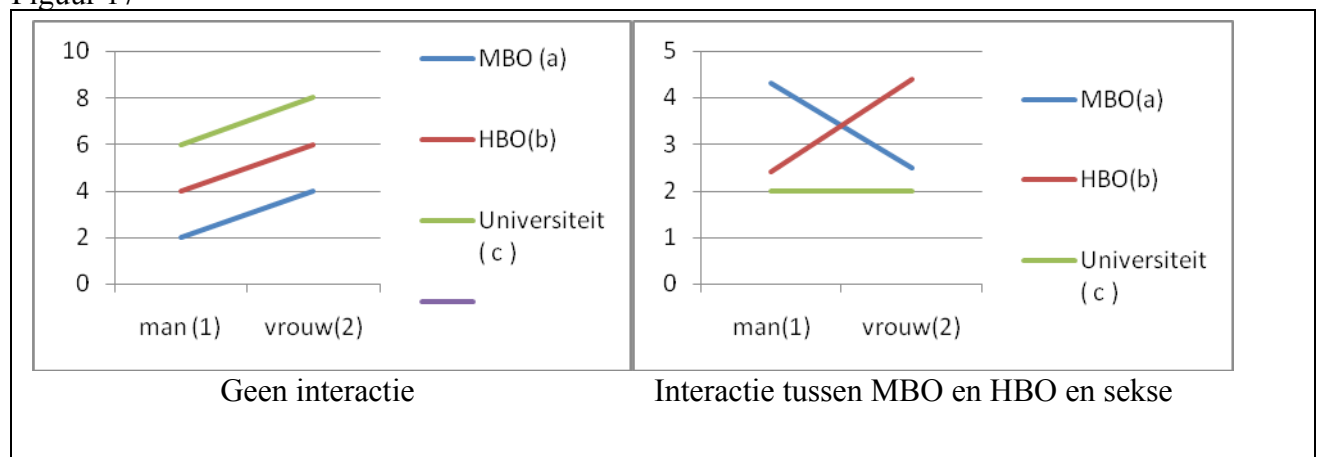
Aanduiding 11: (\bar{y}_2 ; met sample grootte n_2)

Formule 48:

$$\bar{y}_1 - \bar{y}_2 \pm t_{.025} \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

De t score heeft *df*: $N - g$ (totale sample size – aantal groepen)

Figuur 17



Formule 49:

$$F = \frac{\text{Between groups variance van interactie}}{\text{within groups variance}} = \frac{MS \text{ interactie}}{MS \text{ error}}$$