

Formuleblad

Hoofdstuk 1:

Gemiddelde berekenen:

$$\bar{x} = x_1 + x_2 + x_3 + \dots + x_n / n$$

Of: $\bar{x} = 1/n \sum x_i$

Plaats van de median berekenen:

Oneven aantal observaties: $(n+1)/2$

Even aantal observaties: gemiddelde van de twee middelste observaties

Interkwartiele range berekenen:

$$IQR = Q_3 - Q_1$$

Variantie berekenen:

$$s^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 / n - 1$$

Of: $s^2 = 1/n - 1 \sum (x_i - \bar{x})^2$

Standaarddeviatie:

$$\sqrt{s^2}$$

Gestandaardiseerde waarde van x :

$$z = (x - \mu) / \sigma$$

Hoofdstuk 2:

Correlatie berekenen:

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

Formule voor regressielijn :

$$Y = b_0 + b_1 x$$

Waarden b_1 en b_0 vinden met minste voorspellingsfouten:

$$\sum (\text{error})^2 = \sum (y_i - b_0 - b_1 x_i)^2$$

Hoofdstuk 3:

Geen formules

Hoofdstuk 4:

De kans die bij een gebeurtenis hoort:

$$0 \leq P(A) \leq 1$$

Kans op een specifieke uitkomst bij fenomeen met k aantal mogelijke uitkomsten:

$$1/k$$

Optelregel voor disjuncte gebeurtenissen:

$$P(A \text{ of } B) = P(A) + P(B)$$

Vermenigvuldigingsregel voor onafhankelijke gebeurtenissen:

$$P(A \text{ en } B) = P(A)P(B)$$

Gemiddelde van random variabele als a en b vastgestelde getallen zijn:

$$\mu_{a+bX} = a + b\mu_x$$

Gemiddelde van random variabele:

$$\mu_{x+y} = \mu_x + \mu_y$$

Variantie van een random variabele:

$$\sigma_x^2$$

Standaarddeviatie van random variabele als a en b vastgestelde getallen zijn:

$$\sigma_{a+bX}^2 = b^2\sigma_x^2$$

Standaarddeviatie van random variabele:

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$$

$$\text{Ook geldt: } \sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2$$

Standaarddeviatie als X en Y een correlatie ρ hebben:

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y$$

$$\text{Ook geldt: } \sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$$

De kans dat gebeurtenis A en B samen voorkomen (conditionele kans):

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

De kans dat alle gebeurtenissen uit de verzameling voorkomen (intersectie):

$$P(A \text{ en } B \text{ en } C) = P(A) P(B | A) P(C | A \text{ en } B)$$

Hoofdstuk 5 :

Standaarddeviatie van het steekproefgemiddelde:

$$\bar{x} = \sigma / \sqrt{n}$$

De proportie 'successen' in de steekproef bij binomiale data:

$$\hat{p} = X/n$$

Standaarddeviatie van \hat{p} :

$$\sqrt{p(1-p)/n}$$

Aantal manieren waarop k aantal successen bij n aantal observaties geschikt kunnen worden:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Hoofdstuk 6:

Foutenmarge bij het selecteren van SRS van groote n uit populatie met onbekende μ en bekende σ :

$$m = z^*(\sigma/\sqrt{n})$$

Betrouwbaarheidsintervallen:

$$\bar{x} \pm z^* \sigma / \sqrt{n}$$

Hoofdstuk 7:

Standaardfout van het steekproefgemiddelde:

$$SE_M = s / \sqrt{n}$$

Gestandaardiseerde steekproefgemiddelde (één-steekproef-z-toets):

$$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

Één-steekproef-t-toets:

$$t = (\bar{x} - \mu) / (s / \sqrt{n})$$

Betrouwbaarheidsinterval C voor μ :

$$\bar{x} \pm t^*(s / \sqrt{n})$$

T-waarde:

$$t = (\bar{x} - \mu_0) / (s / \sqrt{n})$$

De gepoolde schatter van σ^2 :

$$s_p^2 = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 / (n_1 + n_2 - 2)$$

Gepoolde standaarddeviatie:

$$\sqrt{s_p^2}$$

Gepoolde betrouwbaarheidsinterval:

$$\bar{x}_1 - \bar{x}_2 \pm t^* s_p \sqrt{(s_1^2 / n_1 + s_2^2 / n_2)}$$

F-toets:

$$F = s_1^2 / s_2^2$$

Hoofdstuk 8:

Foutenmarge voor betrouwbaarheidsinterval C:
 $m = z^* SE_{\hat{p}}$

Plus vier regel voor schatten populatieproportie:
 $p\text{-golf} = X + 2 / n + 4$

Standaarddeviatie van de plus vier schatting:
 $\sqrt{p(1-p)/(n+4)}$

Standaardfout van p-golf:
 $\sqrt{p\text{-golf}(1-p\text{-golf})/n+4}$

Foutenmarge van p-golf:
 $m = z^* SE_{p\text{-golf}}$

Vaststellen steekproefgrootte aan de hand van een vaststaande foutenmarge:
 $N = (z^*/m)^2 p^*(1-p^*)$

Steekproefgrootte berekenen na vaststellen van p:*
 $N = 1/4(z^*/m)^2$

Plus vier schattingen van twee populatieproporties:
 $p_1\text{-golf} = X_1 + 1 / n_1 + 2$ en $p_2\text{-golf} = X_2 + 1 / n_2 + 2$

Geschatte verschil tussen populaties:
 $D\text{-golf} = p_1\text{-golf} - p_2\text{-golf}$

Standaardfout van D-golf:
 $\sqrt{(p_1\text{ golf}(1-p_1\text{ golf})/(n_1+2) + p_2\text{ golf}(1-p_2\text{ golf})/(n_2+2))}$

Foutenmarge van D-golf:
 $m = z^* SE_{D\text{-golf}}$

Standaardfout van D bij grote steekproeven:
 $\sqrt{(\hat{P}_1(1-\hat{P}_1)/n_1 + \hat{P}_2(1-\hat{P}_2)/n_2)}$

Foutenmarge voor betrouwbaarheidsinterval van D:
 $m = z^* SE_D$

Hoofdstuk 9:

Verwachte celtelling:
(rijtotaal x kolomtotaal) / n

Chi-kwadraattoets:
 $X^2 = \sum (\text{geobserveerde telling} - \text{verwachte telling})^2 / \text{verwachte telling}$

Hoofdstuk 10:

Regressielijn van de populatie:

$$\mu_y = \beta_0 + \beta_1 x$$

Geobserveerde respons:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Regressielijn voor een steekproef:

$$\hat{y} = b_0 + b_1 x$$

Vinden b_1 en b_0 :

$$b_1 = r(s_y / s_x)$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Residu:

$$e_i = (\text{geobserveerde } y\text{-waarde}) - (\text{voorspelde } y\text{-waarde})$$

Toetsstatistiek voor het testen van H_0 : regressiecoëfficiënt in populatie is 0:

$$t = b_1 / \text{SE}_{b_1}$$

Formules in de ANOVA-tabel :

BRON (source)	Vrijheidsgraden (DF)	SS (sum of squares)	MS (mean square)	F
Model	1	$\Sigma(\hat{y}_i - y\text{-streepje})^2$	SSM/DFM	MSM/MSE
Error	n-2	$\Sigma(y_i - \hat{y}_i)^2$	SSE/DFE	
Totaal	n-1	$\Sigma(y_i - y\text{-streepje})^2$	SST/DFT	

Hoofdstuk 11 :

Formule multipele regressie :

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Statistische model voor multipele lineaire regressie :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

Formules in de ANOVA-tabel voor multipele regressie :

BRON (source)	Vrijheidsgraden (DF)	SS (sum of squares)	MS (mean square)	F
Model	p (aantal predictoren)	$\Sigma(\hat{y}_i - y\text{-streepje})^2$	SSM/DFM	MSM/MSE
Error	n-p-1	$\Sigma(y_i - \hat{y}_i)^2$	SSE/DFE	
Totaal	n-1	$\Sigma(y_i - y\text{-streepje})^2$	SST/DFT	

Verklaarde variantie door alle variabelen samen :

$$R^2 = \text{SSM} / \text{SST}$$

Hoofdstuk 12:

Uitrekenen t^2 :

$$t^2 = n/2(x\text{-gemiddeld} - y\text{-gemiddeld})^2/s_p^2$$

Gepoolde steekproefvariantie :

$$s_p^2 = (n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 + \dots + (n_l - 1) s_l^2 / ((n_1 - 1) + (n_2 - 1) + \dots + (n_l - 1))$$

Gepoolde standaarddeviatie:

$$S_p = \sqrt{s_p^2}$$

Mean Square voor de Error :

$$S_p^2 = \text{MSE} = \text{SSE} / \text{DFE}$$

Formules in de ANOVA-tabel voor eenweg ANOVA :

BRON (source)	Vrijheidsgraden (DF)	SS (sum of squares)	MS (mean square)	F
Groepen	l-1 (aantal groepen-1)	$\sum_{\text{groepen}} n_i(x\text{-gemiddeld}_i - x\text{-gemiddeld})^2$	SSG/DFG	MSG/MSE
Error	N-l	$\sum_{\text{groepen}} (n_i - 1)s_i^2$	SSE/DFE	
Totaal	N-1	$\sum_{\text{obs}} (x_{ij} - x\text{-gemiddeld})^2$		

Coefficiënt van bepaling :

$$R^2 = \text{SSG} / \text{SST}$$

Steekproefcontrast:

$$c = \sum a_i x\text{-gemiddeld}_i$$

Standaardfout van C:

$$SE_c = s_p \sqrt{(\sum a_i^2 / n_i)}$$

T-toetsen om multipale vergelijkingen te toetsen:

$$t_{ij} = x\text{-gemiddeld}_i - x\text{-gemiddeld}_j / (s_p \sqrt{(1/n_i + 1/n_j)})$$

Hoofdstuk 13:

Formules in de tweeweg-ANOVA-tabel :

Bron	Vrijheidsgraden	SS	MS	F
A	I-1	SSA	SSA/DFA	MSA/MSE
B	J-1	SSB	SSB/DFB	MSB/MSE
AB	(I-1)(j-1)	SSAB	SSAB/DFAB	MSAB/MSE
Error	N-IJ	SSE	SSE/DFE	
Totaal	N-1	SST		

Hoofdeffecten toetsen :

$$F_A = MSA/MSE$$

$$F_B = MSB/MSE$$

Interactie-effect toetsen:

$$F_{AB} = MSAB/MSE$$