TMA 4260.

11. Desember 2007.

OPPG 1:

a) $T_A = \frac{Coef}{SECoef} = \frac{1.625}{0.179} \approx 9,08$

SECoef er konstant for alle effekter

$T_B = \frac{0.925}{0.179} \approx 5,17$ // $2 \cdot P(t_4 > 5,17) = 0,007$

$Coef_C = T_C \cdot SECoef = 1,54 \cdot 0,179 \approx 0,275$

$Effekt_C = 2 \cdot Coef_C = 2 \cdot 0,275 \approx 0,55$

$2 \cdot P(t_4 > 1,54) \approx 0,2$

4 frihetsgrader siden 4 effekter brukes til å estimere variansen.
Vi antar at disse er 0, dvs

$AB = AC = BC = ABC = 0$.

$s^2_{effekt} = \frac{1}{4}\left[\hat{AB}^2 + \hat{AC}^2 + \hat{BC}^2 + \hat{ABC}^2\right]$.

forts a)

$$\hat{y}(1,1,1) = \widehat{\text{Constant}} + \frac{\hat{A}}{2} + \frac{\hat{B}}{2} + \frac{\hat{C}}{2}$$

$$= 10{,}05 + 1{,}625 + 0{,}925 + 0{,}275$$

$$\approx \underline{12{,}9}$$

$$\text{Var}[\hat{y}(1,1,1)] = 4 \cdot 0{,}179^2 \approx \underline{0{,}36^2}$$

{4·SECoef² fordi effektene er uavhengige}

95 % konf.int

$$\left[\hat{y}(1,1,1) \pm t_{4,\,0.975} \cdot 0{,}36 \right]$$

$$= \left(12{,}9 - 2{,}78 \cdot 0{,}36 \;,\; 12{,}9 + 2{,}78 \cdot 0{,}36 \right)$$

$$\approx (11{,}9 \;,\; 13{,}9)$$

b)

$$s^2 = \frac{SSE}{6} = \frac{24{,}17}{6} \approx 4{,}03$$

$$\underline{s \approx 2.}$$

forts 6)

$H_0: \alpha_i = 0$ for alle i, $H_1: \alpha_i \neq 0$ for minst en i.

Testobservator

$$F_A = \frac{\frac{SSA}{1}}{\frac{SSE}{6}} \approx \frac{46,4}{2^2} \approx \underline{11,5}$$

Under H_0 er $F_A \sim f_{1,6}$, $f_{1,6,0.95} = \underline{5,99}$.

H_0 forkastes siden $\underline{F_A > 5,99}$.

$H_0: \beta_j = 0$ for alle j, $H_1: \beta_j \neq 0$ for minst en j.

$$F_B = \frac{\frac{SS_B}{2}}{\frac{SSE}{6}} \approx \frac{28.55/2}{2^2} \approx \underline{3.54}$$

Under H_0: $F_B \sim f_{2,6}$, $f_{2,6,0,95} = \underline{5,14}$.

H_0 aksepteres siden $\underline{F_B < 5,14}$.

$H_0: (\alpha\beta)_{ij} = 0$ for alle i,j. $H_1: (\alpha\beta)_{ij} \neq 0$ for minst et i,j par.

$$F_{AB} = \frac{\frac{SS(AB)}{2}}{\frac{SSE}{6}} \approx \frac{1,97/2}{2^2} \approx \underline{0,24}$$

Under H_0: $F_{AB} \sim f_{2,6}$, $f_{2,6,0,95} = \underline{5,14}$

H_0 aksepteres siden $\underline{F_{AB} < 5,14}$.

2 a) $H_0: \mu = 2$, $H_1: \mu \neq 2$

$$T = \frac{\hat{\mu} - 2}{s}\sqrt{18} \sim t_{17} \text{ under } H_0.$$

$$T = \frac{2.97 - 2}{1,37}\sqrt{18} \approx \underline{3,01}$$

$2 \cdot P(t_{17} > 2,898) = 2 \cdot 0,005 = \underline{0,01}$

$2 \cdot P(t_{17} > 3,646) = 2 \cdot 0,001 = \underline{0,002}$.

Her er vi nærmest 0,01, men litt under.

p-verdi $\approx 0,008$.

En ikke-parametrisk test gjør ingen funksjonell-parametrisert antakelse om fordelingen til $y_1, \ldots, y_{18}$.

Fortegnstesten ser på $\#(y_i < 2)$ og $\#(y_i \geq 2)$.

Her er $\#(y_i < 2) = 3$, $\#(y_i > 2) = 15$.

Under $H_0: \mu = 2$ er $\#(y_i < 2) \sim \text{Bin}(18, \frac{1}{2})$.

$P(\#(y_i < 2) \leq 3) \approx 0,004$ { Tabell for binomisk }

p-verdi $= 2 \cdot 0,004 = \underline{0,008}$.

Verdien er veldig lik den som gjør bruk av normalfordeling over.

forts a)

Binomisk kan approksimeres med Normal fra sentralgrenseteoremet.

$$x = \#(y_i < 2) \approx N\left(18 \cdot \frac{1}{2},\ 18 \cdot \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right)\right)$$

$$= N(9,\ 4.5).$$

Da er $\dfrac{x-9}{\sqrt{4.5}} \approx N(0,1)$

H_0 forkastes på signifikansnivå 0,05 dersom

$$\left|\frac{x-9}{\sqrt{4.5}}\right| > z_{0,975} = 1,96$$

b) H_0: $y_1, \ldots, y_{18} \sim N(\mu, \sigma^2)$.

H_1: $y_1, \ldots y_{18}$ er ikke normalfordelte.

Innfører

O_i = observerte antall data i bin i

$i = 1, 2, \ldots, 6$

e_i = forventet antall data i bin i under hypotese H_0, $i = 1, 2, \ldots, 6$.

$$e_1 = 18 \cdot \int_{-\infty}^{0} N(x; 2.97, 1.37^2)\, dx$$

$$= 18 \cdot \Phi\left(\frac{0-2.97}{1.37}\right) \approx \underline{0,27}$$

Her er $\Phi(z)$ kumulativ fordelingsfunksjon for normalfordelingen.

$$e_2 = 18\left[\Phi\left(\frac{1-2.97}{1.37}\right) - \Phi\left(\frac{0-2.97}{1.37}\right)\right] \approx \underline{1,08}$$

$$e_3 = 18\left[\Phi\left(\frac{2-2.97}{1.37}\right) - \Phi\left(\frac{1-2.97}{1.37}\right)\right] \approx \underline{2,95}$$

$$e_4 = 18\left[\Phi\left(\frac{3-2.97}{1.37}\right) - \Phi\left(\frac{2-2.97}{1.37}\right)\right] \approx \underline{4,85}$$

$$e_5 = 18\left[\Phi\left(\frac{4-2.97}{1.37}\right) - \Phi\left(\frac{3-2.97}{1.37}\right)\right] \approx \underline{4,77}$$

$$e_6 = 18\left[\Phi\left(\frac{\infty-2.97}{1.37}\right) - \Phi\left(\frac{4-2.97}{1.37}\right)\right]$$

$$= 18\left[1 - \Phi\left(\frac{4-2.97}{1.37}\right)\right] \approx \underline{4,07}$$

$$X = \sum_{i=1}^{6} \frac{(o_i - e_i)^2}{e_i} \approx \chi^2_{6-1-2} \quad \text{under } H_0$$

Vi mister ytterligere to frihetsgrader pga at μ og σ^2 er estimerte ved $\hat{\mu}$ og s^2.

forts 6)

$$X = \frac{(1-0,27)^2}{0.27} + \frac{(1-1.08)^2}{1.08} + \frac{(1-2.95)^2}{2,95}$$

$$+ \frac{(6-4,85)^2}{4,85} + \frac{(5-4,77)^2}{4,77} + \frac{(4-4,07)^2}{4,07} \approx 3,54.$$

$$\chi^2_{3,\,0.95} = 7,82.$$

H_0 aksepteres siden $X < \chi^2_{3,\,0.95}$.

Selv om histogrammet kan indikere at den sanne underliggende fordelingen er skjev, finner vi ingen grunn til å forkaste hypotesen om normalfordeling utfra vår intervalloppdeling.

3a)

$$s^2 = \frac{SSE}{20-2} = \frac{0,48}{18} \approx \underline{0,16^2}$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{0,48}{24} \approx \underline{98\%}$$

Residualplott:

øverst-venstre: Residual $y_i - \hat{y}_i = e_i$ plottes mot tilpasset normalfordeling. Rett linje indikerer den 'underliggende' normalfordeling. Dersom punktene ligger nær dette har vi cirka normalfordelte residualer.

nederst-venstre: Residual $e_1, \ldots, e_{18}$ plottet som et histogram. Et noksa 'Gauss' formet histogram tyder igjen på at residualene er normalfordelt.

øverst-høyre: Residual $e_1, \ldots, e_{18}$ plottet mot tilpasset verdi $\hat{y}_1, \ldots, \hat{y}_{18}$. Plottet kan evnt. indikere at variansen σ^2 ikke er konstant, men øker med respons.

nederst-høyre: Residual $e_1, \ldots, e_{18}$ plottet mot $1, \ldots, 18$. Plottet kan evnt. indikere at residualene er avhengige av observasjonsrekkefølgen.

I våre plott ser residualplottene noksa fine ut. Ingen spesielle trender som antyder avvik fra modellantakelser.

forts a)

Vi tar her logaritmen for å få en respons på hele tall-linja.

vindstyrke > 0.

$\log(\text{vindstyrke}) \in \mathbb{R}$.

Vi unngår også ekstreme vinder som ville gitt tunge haler, noe som ikke plukkes opp av normalfordelingen.

b) $\hat{\beta}_0 \approx -0{,}217 \qquad \text{Var}[\hat{\beta}_0] \approx 0{,}0999^2$

$$\beta_0 \in (\hat{\beta}_0 \pm t_{18,0,975} \cdot 0{,}099)$$

$$= \underline{(-0{,}427, -0{,}007)}$$

$H_0: \beta_1 = 1, \quad H_1: \beta_1 \neq 1.$

$$T = \frac{\hat{\beta}_1 - 1}{\text{std}(\hat{\beta}_1)} = \frac{1{,}24 - 1}{0{,}042} \approx \underline{5{,}73.}$$

$t_{18,0.975} = 2.1$. H_0 forkastes.

forts b)

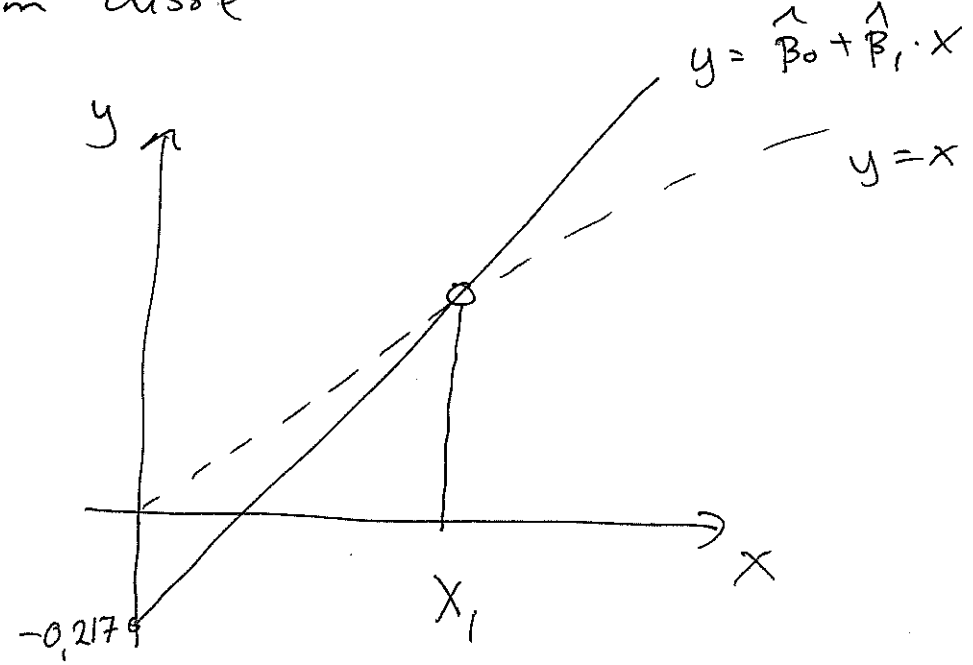
$$E[y] = \beta_0 + \beta_1 \cdot x_1$$

Setter $E[y] = x_1$, Da er

$$x_1 = \beta_0 + \beta_1 \cdot x_1$$

$$x_1 = \frac{\beta_0}{1 - \beta_1}.$$

Denne x tilsvarer en linje der $\beta_0 = 0$, $\beta_1 = 1$, dvs en rett linje gjennom origo.
Da er måling lik varsel.
Når $\beta_0 \neq 0$ og $\beta_1 \neq 1$, finner vi skjæringa mellom disse



Forts b)

$$\hat{X}_1 = \frac{\hat{\beta}_0}{1-\hat{\beta}_1} = \frac{-0,217}{1-1,24} \approx \underline{0,9}$$

$$\hat{X}_1 = g(\hat{\beta}_0, \hat{\beta}_1)$$

$$\approx g(\beta_0, \beta_1) + \frac{\partial g}{\partial \underline{\beta}} \cdot (\hat{\underline{\beta}} - \underline{\beta}) \quad , \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Dvs.

$$\text{var}(\hat{X}_1) \approx \begin{bmatrix} \frac{\partial g}{\partial \hat{\beta}_0} & \frac{\partial g}{\partial \hat{\beta}_1} \end{bmatrix} \cdot \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Kov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Kov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Var}(\hat{\beta}_1) \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial g}{\partial \hat{\beta}_0} \\ \frac{\partial g}{\partial \hat{\beta}_1} \end{bmatrix}$$

$$= \left(\frac{\partial g}{\partial \hat{\beta}_0}\right)^2 \cdot \text{Var}[\hat{\beta}_0] + \left(\frac{\partial g}{\partial \hat{\beta}_1}\right)^2 \cdot \text{Var}[\hat{\beta}_1] + 2 \cdot \frac{\partial g}{\partial \hat{\beta}_0} \frac{\partial g}{\partial \hat{\beta}_1} \cdot \text{Kov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$\approx (4,15)^2 \cdot 0,0999^2 + (3,73)^2 \cdot 0,042^2 - 2 \cdot 4,15 \cdot 3,73 \cdot 0,0037$$

$$\approx \underline{0,22^2}$$

Siden $\frac{\partial g}{\partial \hat{\beta}_0} = \frac{1}{1-\hat{\beta}_1} \approx \frac{1}{1-1,24} \approx \underline{-4,15}$

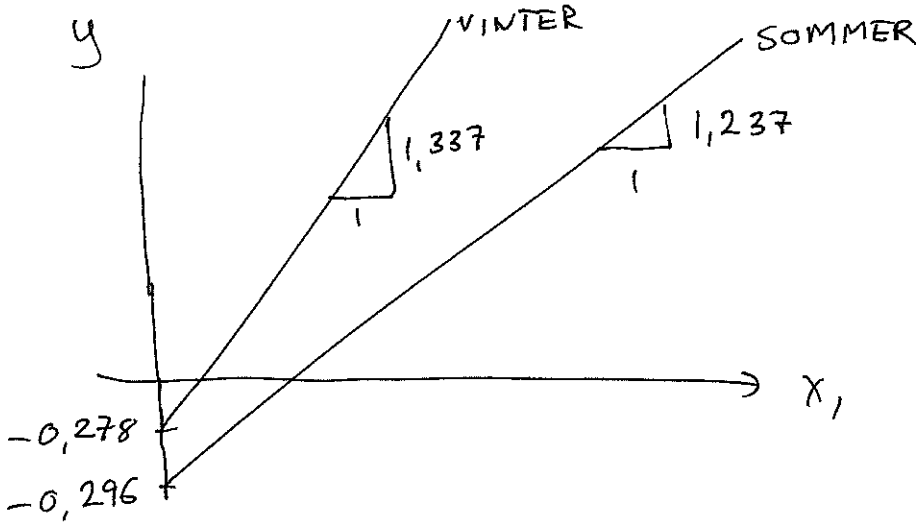
$$\frac{\partial g}{\partial \hat{\beta}_1} = \frac{\hat{\beta}_0}{(1-\hat{\beta}_1)^2} = \frac{-0,217}{(1-1,24)^2} \approx \underline{-3,73}$$

c) Her blir

$$\hat{y}_{SOMMER} = -0,296 + 1,237 \cdot X_1$$

$$\hat{y}_{VINTER} = -0,296 + 1,237 \cdot X_1 + 0,018 \cdot 1 + 0,10 \cdot X_1 \cdot 1$$

$$= -0,278 + 1,337 \cdot X_1$$



F-test:

$$F = \frac{[SSR(4\ par) - SSR(2\ par)]/2}{SSE(4\ par)/16}$$

$$\approx \frac{(23,8 - 23.52)/2}{0,2/16} \approx 11,2$$

Under H_0: β_2 og β_3 lik 0 er

$F \sim f_{2,16}$. $f_{2,16,\ 0.95} = 3.6$.

H_0 forkastes.

d)

$$\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_{20} \\ x_{1,1} \\ \vdots \\ x_{20,1} \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \\ 0 & & & 1 & 0 \\ 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \\ 0 & & & 1 & 1 \end{bmatrix}}_{\underline{X}} \begin{pmatrix} z_1 \\ \vdots \\ z_{20} \\ \beta_0 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{20} \\ \delta_1 \\ \vdots \\ \delta_{20} \end{pmatrix}$$

$$\mathrm{Var}[v_1] = \cdots = \mathrm{Var}[v_{20}] = \sigma^2, \quad \text{pga } \varepsilon_1, \ldots, \varepsilon_{20}.$$

$$\mathrm{Var}[v_{21}] = \cdots = \mathrm{Var}[v_{40}] = \tau^2, \quad \text{pga } \delta_1, \ldots, \delta_{20}.$$

Minste kvadraters estimator minimerer altså

$$\sum_i \left(y_i - (\underline{X} \cdot \underline{\beta})_i\right)^2 = (\underline{y} - \underline{X} \cdot \underline{\beta})^T \cdot (\underline{y} - \underline{X} \cdot \underline{\beta}).$$

Her må de ulike målingene veies mhp τ^2 og σ^2.

Vi minimerer

$$(\underline{y} - \underline{X} \cdot \underline{\beta})^T \cdot W \cdot (\underline{y} - \underline{X} \cdot \underline{\beta}) = \ell(\underline{\beta}).$$

Dette er identisk med maksimum-likelihood for $\underline{\beta}$ i denne modellen.

forts d)

$$\frac{\partial l}{\partial \underline{\beta}} = \frac{\partial}{\partial \underline{\beta}} \left\{ \underline{y}' W \underline{y} - 2 \underline{\beta}' X' W \underline{y} + \underline{\beta}' X' W X \underline{\beta} \right\}$$

$$= -2 X' W \underline{y} + 2 X' W X \underline{\beta}$$

Løsning er $\frac{\partial l}{\partial \underline{\beta}} = 0$.

Dvs

$$X' W \underline{y} = X' W X \cdot \hat{\underline{\beta}}$$

$$\underline{\hat{\beta}} = [X' W X]^{-1} X' W \underline{y}$$

$$\text{Var}[\hat{\underline{\beta}}] = [X' W X]^{-1} X' W W^{-1} W X [X' W X]^{-1}$$

$$= [X' W X]^{-1}$$

Dersom $W = \sigma^{-2} I$ blir uttrykkene

$$\hat{\underline{\beta}} = [X' \sigma^{-2} I X]^{-1} X' \sigma^{-2} I \underline{y}$$

$$= [X' X]^{-1} X' \underline{y}.$$

$$\text{Var}[\hat{\underline{\beta}}] = [X' \sigma^{-2} I X]^{-1} = \sigma^2 [X' X]^{-1}$$