ST2202 Des-07

1 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.

$\underline{\text{p-verdi} \approx 0,008}$.

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

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

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

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

forts 1a)

Binomisk kan approksimeres med Normal fra sentralgrenseteoremet.

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

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

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

forts 1b)

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

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

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

$$e_2 = 18\left[\Phi\left(\frac{1-2{,}97}{1.37}\right) - \Phi\left(\frac{0-2{,}97}{1.37}\right)\right] \approx \underline{1{,}08}$$

$$e_3 = 18\left[\Phi\left(\frac{2-2{,}97}{1.37}\right) - \Phi\left(\frac{1-2{,}97}{1.37}\right)\right] \approx \underline{2{,}95}$$

$$e_4 = 18\left[\Phi\left(\frac{3-2{,}97}{1.37}\right) - \Phi\left(\frac{2-2.97}{1.37}\right)\right] \approx \underline{4{,}85}$$

$$e_5 = 18\left[\Phi\left(\frac{4-2{,}97}{1.37}\right) - \Phi\left(\frac{3-2{,}97}{1.37}\right)\right] \approx \underline{4{,}77}$$

$$e_6 = 18\left[\Phi\left(\frac{\infty-2{,}97}{1.37}\right) - \Phi\left(\frac{4-2.97}{1.37}\right)\right]$$

$$= 18\left[1 - \Phi\left(\frac{4-2{,}97}{1.37}\right)\right] \approx \underline{4{,}07}$$

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

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

forts 1b)

$$X = \frac{(1-0{,}27)^2}{0{,}27} + \frac{(1-1{,}08)^2}{1{,}08} + \frac{(1-2{,}95)^2}{2{,}95}$$

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

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

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

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.

2a)

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

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

Residualplott:

<u>øverst-venstre:</u> 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.

<u>nederst-venstre:</u> Residual $e_1, \dots, e_{18}$ plottet som et histogram. Et nokså 'Gauss' formet histogram tyder igjen på at residualene er normalfordelt.

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

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

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

forts 2a)

Vi tar her logaritmen for å få en respons på hele tall-linja.
vindstyrke > 0.
log(vindstyrke) $\in \mathbb{R}$.

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

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

$\beta_0 \in (\hat{\beta}_0 \pm t_{18,0,975} \cdot 0,0999)$

$= (-0,427, -0,007)$

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

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

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

6.

forts 2b)

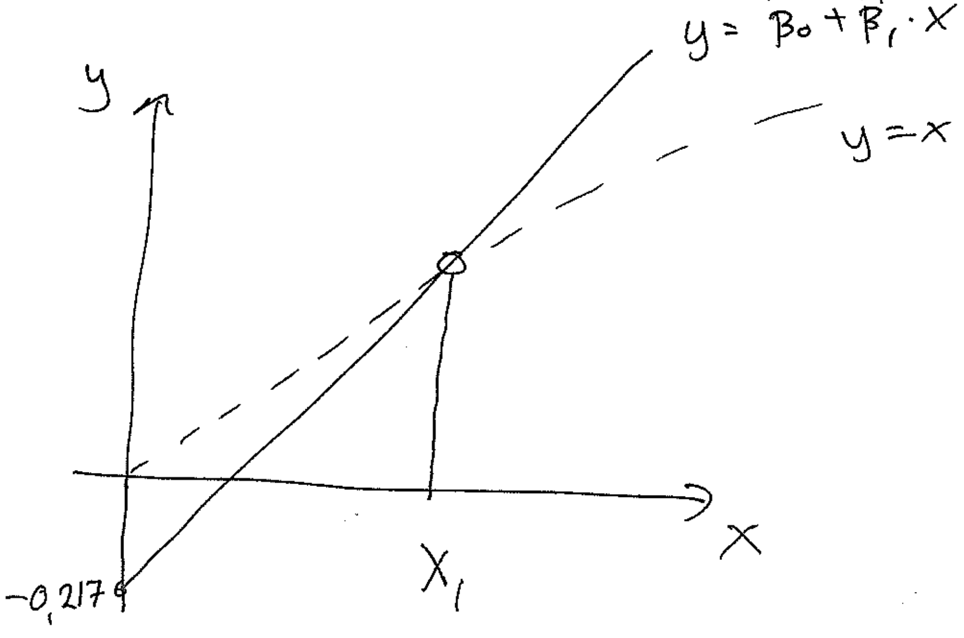
$$E[y] = \beta_0 + \beta_1 \cdot X_1$$

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 2b)

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

$$\hat{X}_1 = g(\hat{\beta}_0, \hat{\beta}_1)$$
$$\approx g(\beta_0, \beta_1) + \frac{\partial g}{\partial \underline{\beta}} \cdot (\hat{\underline{\beta}} - \underline{\beta}) \quad , \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Dvs.

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

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

$$\approx (4,15)^2 \cdot 0,0999^2 + (3.73)^2 \cdot 0,042^2 - 2 \cdot 4,15 \cdot 3.73 \cdot 0,0037$$

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

Siden
$$\frac{\partial g}{\partial \hat{\beta}_0} = \frac{1}{1-\hat{\beta}_1} \approx \frac{1}{1-1,24} \approx \underline{-4,15}$$
$$\frac{\partial g}{\partial \hat{\beta}_1} = \frac{\hat{\beta}_0}{(1-\hat{\beta}_1)^2} = \frac{-0,217}{(1-1,24)^2} \approx \underline{-3,73}$$

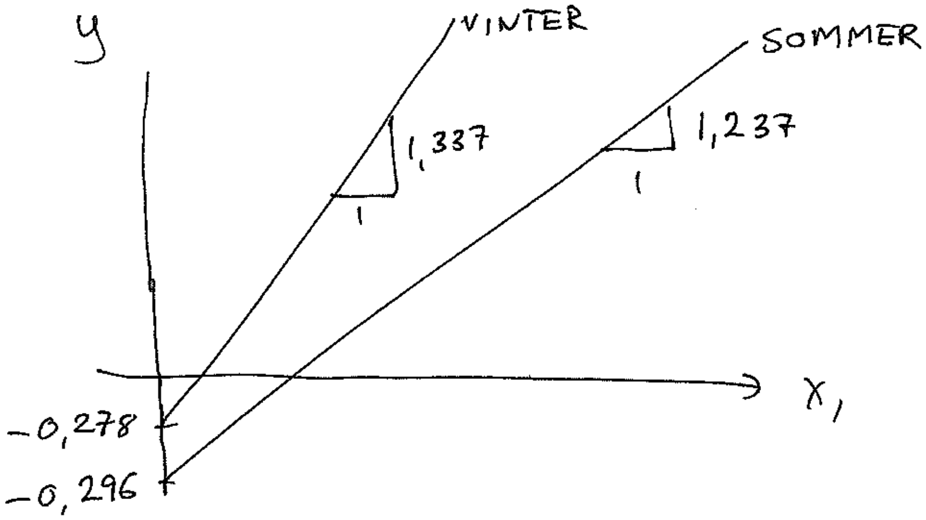
2c)

Her blir

$$\underline{\hat{y}^{\text{SOMMER}} = -0{,}296 + 1{,}237 \cdot X_1.}$$

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

$$= \underline{-0{,}278 + 1{,}337 \cdot X_1.}$$



F-test:

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

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

Under H_0: β_2 og β_3 lik 0 er

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

<u>H_0 forkastes.</u>

2d)

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

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

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

Minste kvadraters estimator minimerer oftest

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

Her må de ulike målingene vektes mtp 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 2d)

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

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

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

Dvs

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

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

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

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

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

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

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

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

3a)

- Data y_i, $i=1,\dots m$ er binomisk fordelt $p(y_i) = \binom{n_i}{y_i} p^{y_i} (1-p)^{n_i - y_i}$.

- $$p = p(\underline{x}_i) = \frac{e^{\beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k}}}{1 + e^{\beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k}}}$$

der $\underline{x}_i$ er kjente kovariater, $i=1,\dots,m$.

Her er $m = 12$, $k = 2$ med kovariater

$x_{i,1}$ = Setninger
$x_{i,2}$ = formel.

$$Z = \frac{\text{Coef}}{\text{SECoef}} = \frac{2.64}{1,05} \approx \underline{\underline{2,51}}$$

$$\text{SECoef} = \frac{\text{Coef}}{Z} = \frac{1,38}{2,36} \approx \underline{\underline{0,58}}$$

$$\text{Coef} = Z \cdot \text{SECoef} = -2.93 \cdot 0.18 = \underline{\underline{-0,54}}$$

$$p = 2 \cdot P(Z < -2.93) = 2 \cdot 0,0017 \approx \underline{\underline{0,003}}$$

Alle p-verdier er små, og dermed bør alle kovariatene være med i modellen.

$$p(1,6,1) = \frac{1}{1 + e^{-2.64 - 1.38 + 0.54 \cdot 6}} \approx \underline{\underline{0,686}}$$

3 b) $o(x) = \frac{p(x)}{1-p(x)} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$

$$\hat{o}(1,4,1) = e^{\hat{\beta}_0 + \hat{\beta}_1 \cdot 4 + \hat{\beta}_2 \cdot 1}$$
$$\hat{o}(1,4,0) = e^{\hat{\beta}_0 + \hat{\beta}_1 \cdot 4 + \hat{\beta}_2 \cdot 0}$$

$$\frac{\hat{o}(1,4,1)}{\hat{o}(1,4,0)} = e^{\hat{\beta}_2}$$

dvs $\underline{\underline{\hat{o}(1,4,1) = \hat{o}(1,4,0) \cdot e^{\hat{\beta}_2}}}$

$$e^{\hat{\beta}_2} = e^{1,38} \approx \underline{\underline{3,97}}$$

$$\mathrm{Var}[e^{\hat{\beta}_2}] = \mathrm{Var}[g(\hat{\beta}_2)]$$
$$\approx \left(\frac{\partial g}{\partial \hat{\beta}_2}\right)^2 \cdot \mathrm{Var}[\hat{\beta}_2]$$
$$= \left(e^{\hat{\beta}_2}\right)^2 \cdot 0,58^2$$
$$\approx (3,97)^2 \cdot 0,58^2$$
$$\approx \underline{\underline{2,3^2}}$$

<u>95% konf int:</u>

$$e^{\beta_2} \in \left\{ e^{\hat{\beta}_2} \pm 1,96 \cdot \sqrt{\widehat{\mathrm{Var}}(e^{\hat{\beta}_2})} \right\} = \underline{(-0,54,\ 8,48)}$$

13.

3 c) vår estimator gir

$$\hat{r}_1 = \left[\frac{\hat{o}(1,4,1)}{\hat{o}(1,4,0)}\right] = \underline{3,97}$$

$$\underline{\mathrm{var}[\hat{r}_1] = 2,3^2.}$$

Annen person får: $\underline{\hat{r}_2 = 3,0}$, $\underline{\mathrm{var}[\hat{r}_2] = 1,0^2}$

Vi kombinerer disse for optimalt estimat:

$$\hat{r} = a \cdot \hat{r}_1 + (1-a) \cdot \hat{r}_2$$

best valg av a, som minimerer $\mathrm{Var}[\hat{r}]$

er $a = \dfrac{\mathrm{Var}[\hat{r}_2]}{\mathrm{var}[\hat{r}_1] + \mathrm{Var}[\hat{r}_2]}$

$$\hat{r} = \frac{1,0^2}{1,0^2 + 2,3^2} \cdot 3,97 + \frac{2,3^2}{1,0^2 + 2,3^2} \cdot 3,0 \approx \underline{\underline{3,16}}$$

$$\mathrm{Var}[\hat{r}] = a^2 \cdot \mathrm{Var}[\hat{r}_1] + (1-a)^2 \cdot \mathrm{Var}[\hat{r}_2]$$

$$= 0,16^2 \cdot 2,3^2 + 0.84^2 \cdot 1,0^2$$

$$\underline{\underline{= 0,84 = 0.92^2}}$$

3 d).

$\ell(3 \text{ parametre}) = -39{,}4$

$\ell(2 \text{ parametre}) = -42{,}3$

$$D = 2 \cdot [\ell(3 \text{ parametre}) - \ell(2 \text{ parametre})]$$
$$= 2 \cdot [-39{,}4 + 42.3] = \underline{5{,}8}$$

Under H_0: 'formel' $= 0$, H_1: 'formel' $\neq 0$

er $D \approx \chi^2_{3-2} = \chi^2_1$.

$\underline{\chi^2_{1,0.95} = 3{,}84}$, $\underline{\chi^2_{1,0.975} = 5{,}02}$, $\underline{\chi^2_{1,0.99} = 6{,}64}$

<u>p-verdi $\approx 0{,}02$.</u>

I punkt a) fikk vi p-verdi marginalt for formel på 0,018.
Disse er nokså like her.