LØSNINGSFORSLAG, EKSAMEN DES' 06

OPPG 1: a) $T = \frac{\text{Coef}}{\text{SECoef}} = \frac{\hat{\beta}_i}{\text{Std}(\hat{\beta}_i)}$

$\text{SECoef} = \frac{105,58}{3,29} \approx 16,80$

$T = -35.58/9,58 \approx -3,71$

$\text{Coef} = 56,85 \cdot 1 = 56,85$

$s^2 = \frac{SSE}{n-(k+1)} = \frac{2841,9}{9-4} = 23,8^2$

Residual: $e_i = y_i - \hat{y}_i$, $i = 1, \ldots, 9$.

Forventer at e_i "oppfører seg" som ε_i, dvs uavhengige, konstant varians, normalfordelte. En 'trend' som detektes i residualplottene her er at variansen σ^2 muligens øker med stigende $\hat{y}$ (se Figur øverst til høyre).

$R^2 = \frac{SSR}{SST} = \frac{8422}{11264} \approx 0,75$

$SSR = \sum_{i=1}^{9} (\hat{y}_i - \bar{y})^2$, $SST = \sum_{i=1}^{9} (y_i - \bar{y})^2$.

1b) $H_0: \beta_2 = \beta_3 = 0$, $H_1:$ minst en av $\beta_2, \beta_3 \neq 0$.

Testes ved å se på $SSR(\beta_0, \beta_1)$ versus $SSR(\beta_0, \beta_1, \beta_2, \beta_3)$. Disse oppgis i Seq SS.

Testobservator:

$$F = \frac{[SSR(\beta_0, \beta_1, \beta_2, \beta_3) - SSR(\beta_0, \beta_1)]/2}{SSE/9-4}$$

$$= \frac{[8422,1 - 7532,4]/2}{23,8^2} = \frac{(0.1 + 889,7)/2}{23,8^2}$$

$$\simeq 0,78$$

H_0 forkastes dersom $F > f_{2,5,0.95} = 5,79$.

Konklusjon:
H_0 forkastes ikke.

Tolkningen blir da at nedbør for detober kan forklares v.h.a. et konstantledd og østkoordinat. De ekstra faktorvariable, nord og høyde forklarer ikke merkebart mer av variasjonen i datamaterialet.

1c. $y_0 = \beta_0 + \beta_1 x_{1,0} + \varepsilon_0$

$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_{1,0} = 124 + 24{,}8 \cdot 0.75 \approx 146{,}35$

$E[\hat{y}_0 - y_0] = 0$

$$Var[\hat{y}_0 - y_0] = [1, x_{1,0}] \begin{bmatrix} \sigma^2 0.111 & \sigma^2 0{,}001 \\ \sigma^2 0.001 & \sigma^2 0.117 \end{bmatrix} \begin{bmatrix} 1 \\ x_{1,0} \end{bmatrix} + \sigma^2$$

$$= \sigma^2 (0.111 + 0.75^2 \cdot 0{,}117 + 2 \cdot 0.75 \cdot 0.001 + 1)$$

$$= \sigma^2 \cdot 1.17$$

σ^2 estimeres vha $s^2 = 23{,}09^2$. $n-2 = 7$ frihetsgrader

$$\frac{\hat{y}_0 - y_0 - 0}{s\sqrt{1{,}17}} \sim t_7$$

95% intervall

$$y_0 \in \left(\hat{y}_0 \pm t_{7,0.975} \cdot s \cdot \sqrt{1{,}17}\right) = (87.2, 205.5)$$

4,75° øst for Trondheim er 15.15° E og langt utenfor området der vi har data.
En slik prediksjon vil dermed trolig ikke gjelde.

OPPG 2:

a) Generator $D = ABC$.

Konfunderingsmønster

$D \sim ABC$, $C \sim ABD$, $B \sim ACD$, $A \sim BCD$

$AB \sim AD$, $AC \sim BD$, $AD \sim BC$

Klarer ikke å skille disse.

$$\hat{C} = \frac{3.77 + 3.03 + 4.44 + 5.25}{4} - \frac{3.62 + 4.01 + 3.23 + 4.93}{4}$$

$$= \underline{0.175}$$

AB kolonnen er $\begin{pmatrix} + \\ - \\ + \\ + \\ - \\ - \end{pmatrix}$

$$\hat{AB} = \frac{3.77 + 4.44 + 3.62 + 4.01}{4} - \frac{3.03 + 3.23 + 5.25 + 4.93}{4}$$

$$= \underline{-0.15}$$

b) Bruker $\hat{AB}, \hat{AC}, \hat{AD}$ til å estimere Var[effekt]

$$s^2_{\hat{effekt}} = \frac{1}{3}\{\hat{AB}^2 + \hat{AC}^2 + \hat{AD}^2\}$$

$$= \frac{1}{3}\{0{,}15^2 + 0{,}14^2 + 0.4^2\} = \underline{0{,}258^2}$$

$$\frac{\hat{Effekt} - Effekt}{s_{effekt}} \sim t_3$$

H_0: Effekt $= 0$, H_1: Effekt $\neq 0$

H_0 forkastes dersom $|\hat{Effekt}| > t_{3,0.975} \cdot 0{,}258$

$= 3.182 \cdot 0.258 = \underline{0{,}82}$

Dvs A og B er signifikante.

$$\hat{y}_{-+} - \hat{y}_{++} = \left(\hat{\mu} - \frac{\hat{A}}{2} + \frac{\hat{B}}{2}\right) - \left(\hat{\mu} + \frac{\hat{A}}{2} + \frac{\hat{B}}{2}\right) = -\hat{A}$$

$$\frac{-\hat{A} + A}{s_{eff}} \sim t_3$$

95% konf.int $= \{-\hat{A} \pm t_{3,0.975} \cdot 0{,}258\}$

$= \underline{(-1{,}73 \ , \ -0.09)}$

OPPG 3

3a) $SSA = SST - SSE = 250,9 - 237,8 \simeq 13,1$

$MSA = SSA/2 = \frac{13.1}{2} = 6,55$ $\quad MSE = 237.8/9 \simeq 26,4$

$DF_{error} = N-k = 12-3 = 9$, $\quad DF_{group} = 3-1 = k-1 = 2$

$$F = \frac{MSA}{MSE} = \frac{6.05}{26.4} \simeq 0.23.$$

$H_0: \mu_1 = \mu_2 = \mu_3$, $\quad H_1: \mu_i \neq \mu_j$ for minst ett (i,j) par.

$f_{2,9,0.95} = 4,26$. H_0 forkastes dersom $F > 4.26$.
Her forkaster H_0 ikke!

3b) Normalantakelsen sjekkes vha histogram. De skal danne søyler a la normaltetthetsen. Alternativt kan man benytte Goodness-of-fit tester der utfallsrommet deles inn i intervall og man sammenligner med forventet andel utfra normaltetthet. En tredje muligheter er Q-Q plot eller normal-normal, der data sorteres og plottes med normalpercentiler.

Kruskal-Wallis-test: $H = \frac{12}{N(N+1)} \sum_{i=1}^{3} \frac{r_i^2}{n_i} - 3(N+1) \approx \chi^2_2$

$n_1 = n_2 = n_3 = 4$, $\quad N = 12$.

Gruppe	3	1	3	2	1	2	2	1	3	3	2	1
Rang	1	2	3	4	5	6	7	8.5	8.5	10	11	12
Obs	17	18	20	22	23	24	25	26	26	29	31	32

Dette gir $r_1 = 2 + 5 + 8.5 + 12 = 27.5$

$r_2 = 4 + 6 + 7 + 11 = 28$

$r_3 = 1 + 3 + 8.5 + 10 = 22.5$

$$H = \frac{12}{12 \cdot 13} \sum_{i=1}^{3} \frac{r_i^2}{n_i} - 3 \cdot 13$$

$$= \frac{1}{13}\left[\frac{27.5^2}{4} + \frac{28^2}{4} + \frac{22.5^2}{4}\right] - 39$$

$$= \underline{0.36}$$

$$\chi^2_{2,0.95} = 5.99.$$

H_0 forkastes ikke.

Samme konklusjon som i 1a)

c)
$$y = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -\frac{1}{2}\sqrt{\frac{2}{3}} & -\frac{1}{2}\sqrt{\frac{2}{3}} & \sqrt{\frac{2}{3}} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Her er $q_1' \cdot q_1 = 1$, $q_1' \cdot q_2 = 0$, $q_1' \cdot q_3 = 0$

$q_2' \cdot q_2 = 1$, $q_2' \cdot q_3 = 0$, $q_3' \cdot q_3 = 1$

dvs. $A' \cdot A = I$

$$q_3' \cdot q_2 = -\frac{1}{2}\sqrt{\frac{2}{3}}\left(-\frac{1}{\sqrt{2}}\right) - \frac{1}{2}\sqrt{\frac{2}{3}}\left(\frac{1}{\sqrt{2}}\right) = 0$$

$$q_3' \cdot q_1 = -\frac{1}{2}\frac{1}{\sqrt{3}}\sqrt{\frac{2}{3}} - \frac{1}{2}\sqrt{\frac{1}{3}}\sqrt{\frac{2}{3}} + \frac{1}{\sqrt{3}}\sqrt{\frac{2}{3}} = 0$$

$$q_3' \cdot q_3 = \frac{1}{4} \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{2}{3} + \frac{2}{3} = 1.$$

$$\underline{q}_1' \cdot \underline{q}_2 = \frac{1}{\sqrt{6}}\left(-\frac{1}{\sqrt{2}}\right) + \frac{1}{\sqrt{3}}\left(\frac{1}{\sqrt{2}}\right) = 0$$

$$\underline{q}_1' \cdot \underline{q}_1 = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = \underline{1}.$$

$$\underline{q}_2' \cdot \underline{q}_2 = \frac{1}{2} + \frac{1}{2} = \underline{1}$$

$$E[y_1] = \frac{1}{\sqrt{3}}\left\{E[x_1] + E[x_2] + E[x_3]\right\} = \underline{\sqrt{3}\cdot\mu}.$$

$$E[y_2] = -\frac{1}{\sqrt{2}}E[x_1] + \frac{1}{\sqrt{2}}E[x_2] = \frac{-1}{\sqrt{2}}\mu + \frac{1}{\sqrt{2}}\mu = \underline{0}$$

$$E[y_3] = -\frac{1}{2}\sqrt{\frac{2}{3}}E[x_1] - \frac{1}{2}\sqrt{\frac{2}{3}}E[x_2] + \sqrt{\frac{2}{3}}E[x_3]$$

$$= -\sqrt{\frac{2}{3}}\mu + \sqrt{\frac{2}{3}}\mu = \underline{0}$$

$$Var[y] = A \cdot Var(x) \cdot A' = A \cdot \sigma^2 \cdot I \cdot A' = \sigma^2 A \cdot A'$$

$$= \underline{\sigma^2 \cdot I}$$

y_1, y_2, y_3 er uavhengige.
De er en lineærkombinasjone av x og dermed Normalfordelte.

Metode I for regneglykke:

$$\sum_{i=1}^{3}(x_i-\bar{x})^2 = \sum_{i=1}^{3}x_i^2 - 3\cdot\bar{x}^2 = \sum_{i=1}^{3}x_i^2 - y_1^2$$

Dersom $\sum_{i=1}^{3}y_i^2 = \sum_{j=1}^{3}x_i^2$ gå dette opp.

$$\sum_{i=1}^{3}x_i^2 = x'\cdot x \quad , \quad \sum_{i=1}^{3}y_i^2 = y'\cdot y$$

$$y'\cdot y = [A\cdot x]'\cdot[A\cdot x] = x'\cdot A'\cdot A\cdot x = x'\cdot I\cdot x = x'\cdot x.$$

$$\frac{y_2}{\sigma}\sim N(0,1) \qquad \frac{y_3}{\sigma}\sim N(0,1)$$

$$\frac{y_2^2}{\sigma^2}\sim \chi_1^2 \qquad \frac{y_3^2}{\sigma^2}\sim \chi_1^2$$

$$\frac{y_2^2}{\sigma^2}+\frac{y_3^2}{\sigma^2}\sim \chi_2^2$$

siden frihetsgradene summeres.

Metode 2 for mellomregning:

$$y_1^2 = 3\cdot\bar{x}^2 = 3\cdot\frac{(x_1+x_2+x_3)^2}{3^2} = \frac{1}{3}\left(x_1^2+x_2^2+x_3^2+2x_1x_2+2x_1x_3+2x_2x_3\right)$$

$$y_2^2 = \frac{1}{2}(x_2-x_1)^2 = \frac{1}{2}\left(x_2^2-2x_2x_1+x_1^2\right)$$

$$y_3^2 = \frac{2}{3}\left(x_3-\frac{x_2}{2}-\frac{x_1}{2}\right)^2 = \frac{1}{6}\left[2x_3-x_2-x_1\right]^2$$

$$= \frac{1}{6}\left(4x_3^2+x_2^2+x_1^2-4x_3x_2-4x_3x_1+2x_2x_1\right)$$

Samler ledd:

$$y_1^2+y_2^2+y_3^2 = x_1^2\left[\frac{1}{3}+\frac{1}{2}+\frac{1}{6}\right]+x_2^2\left[\frac{1}{3}+\frac{1}{2}+\frac{1}{6}\right]$$

$$+x_3^2\left[\frac{1}{3}+\frac{4}{6}\right]+x_1x_2\left[\frac{2}{3}-\frac{2}{2}+\frac{2}{6}\right]$$

$$+x_1x_3\left[\frac{2}{3}-\frac{4}{6}\right]+x_2x_3\left[\frac{2}{3}-\frac{4}{6}\right]$$

$$= x_1^2+x_2^2+x_3^2$$