Fasit eksamen STK1110-h06

Oppgave 1

a) Se Rice side 248 og 252.

b) Fra formelsamling eller pensum: $E(\hat{p}) = p, Var(\hat{p}) = p(1-p)/n$. Ved Chebysjevs's ulikhet $P(|\hat{p}-p| > \epsilon) = P(|\hat{p} - E(\hat{p}| > \epsilon) =< Var(\hat{p})/\epsilon^2 = p(1-p)/(n\epsilon^2) \to 0$ når $n \to \infty$.

c) $E(\hat{\theta}) = E(\hat{p}) - E(\hat{p})^2 = E(\hat{p}) - (Var(\hat{p}) + E(\hat{p})^2) = (n-1)p(1-p)/n = (n-1)\theta/n$. La $\tilde{\theta} = n\hat{\theta}/(n-1)$. Da er $E(\tilde{\theta}) = p(1-p) = \theta$.

Oppgave 2

a)

$$lik(\sigma) = \prod_{i=1}^{n} \frac{1}{2\sigma} \exp(\frac{1}{\sigma}\sum_{i=1}^{n}|x_i|), l(\sigma) = n\log(2) - n\log(\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n}|x_i|,$$

$$l'(\sigma) = -n/\sigma + \frac{1}{\sigma^2}\sum_{i=1}^{n}|x_i|, l'(\sigma)|_{\sigma=\hat{\sigma}} = 0 \Rightarrow \hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}|x_i|.$$

$l''(\hat{\sigma}) < 0$ slik at $\hat{\sigma} = \sum_{i=1}^{n}|x_i|$ er SME. $E(|X|) = \int_{-\infty}^{\infty}|x|\frac{1}{2\sigma}\exp(-\frac{1}{\sigma}|x|)dx = \int_0^\infty x\frac{1}{\sigma}\exp(-\frac{1}{\sigma}x)dx$. Forventning i gammafordeling med formparameter 1 og skalaparameter $\frac{1}{\sigma}$ er $1/(1/\sigma) = \sigma$. Herav $E(\hat{\sigma}) = E(|X|) = \sigma$.

b) Fra pensum $\sqrt{nI(\sigma)}(\hat{\sigma}-\sigma)$ tilnærmet $N(0,1)$-fordelt med $I(\sigma) = -E[\frac{\partial^2}{\partial\sigma^2}\log f(X|\sigma)] = -E[\frac{1}{\sigma^2} - \frac{2|X|}{\sigma^3}] = \frac{1}{\sigma^2}$. Herav $\sqrt{n}(\frac{\hat{\sigma}-\sigma}{\sigma})$ tilnærmet $N(0,1)$-fordelt.

c) Også $\sqrt{n}(\frac{\hat{\sigma}-\sigma}{\hat{\sigma}})$ er tilnærmet $N(0,1)$-fordelt slik at

$$1-\alpha \quad \approx \quad P(-z_{1-\alpha/2} < \sqrt{n}(\frac{\hat{\sigma}-\sigma}{\hat{\sigma}}) < z_{1-\alpha/2}) = P(\hat{\sigma} - z_{1-\alpha/2}\hat{\sigma}/\sqrt{n} < \sigma < \hat{\sigma} + z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}).$$

Viste i punkt b) at $\sqrt{n}(\frac{\hat{\sigma}-\sigma}{\sigma})$ er tilnærmet $N(0,1)$-fordelt slik at

$$\begin{aligned}
1-\alpha &\approx P(-z_{1-\alpha/2} < \sqrt{n}(\frac{\hat{\sigma}-\sigma}{\sigma}) < z_{1-\alpha/2}) \\
&= P(-z_{1-\alpha/2}\sigma/\sqrt{n} < (\hat{\sigma}-\sigma) < z_{1-\alpha/2}\sigma/\sqrt{n}) \\
&= P(\sigma(1 - z_{1-\alpha/2}/\sqrt{n}) < \hat{\sigma} < \sigma(1 + z_{1-\alpha/2}/\sqrt{n})) \\
&= P(\frac{\hat{\sigma}}{1 + z_{1-\alpha/2}/\sqrt{n}} < \sigma < \frac{\hat{\sigma}}{1 - z_{1-\alpha/2}/\sqrt{n}})
\end{aligned}$$

Oppgave 3

a) $\hat{\beta}_1 = \frac{\sum x_i y_i - (\sum x_j)(\sum y_j)/n}{\sum x_i^2 - (\sum x_j)^2/n} = \frac{8653.45 - 511.5 \times 267.1/17}{16628.7 - (511.5)^2/17} = 0.4981$. $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 267.1/17 - 511.51 \times 0.4981/17 = 0.725$.

b) Minimerer $\sum (y_i - \gamma x_i)^2$ med hensyn på γ. Løsning $\hat{\gamma} = \frac{\sum x_i y_i}{\sum x_i^2}$ Insatt $\hat{\gamma} = 8653.45/16628.7 = 0.52$

c) $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} = \sum (\frac{1}{n} + \frac{x_i - \bar{x}}{\sum (x_j - \bar{x})^2} \bar{x}) Y_i$ er en lineærkombinasjon av normalfordelte variable og derfor også normalfordelt. Fra pensum eller formelsamliming $E(\hat{\beta}_0) = \beta_0$ og $Var(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n(\sum x_i^2 - n\bar{x}^2)}$. Dette viser at

$$\frac{\sqrt{n}(\hat{\beta}_0 - \beta_0)}{\sigma \sqrt{\frac{\sum x_i^2}{\sum x_i^2 - n(\bar{x})^2}}}$$

er $N(0,1)$-fordelt og at

$$\frac{\sqrt{n}(\hat{\beta}_0 - \beta_0)}{S \sqrt{\frac{\sum x_i^2}{\sum x_i^2 - n(\bar{x})^2}}} = \frac{\frac{\sqrt{n}(\hat{\beta}_0 - \beta_0)}{\sigma \sqrt{\frac{\sum x_i^2}{\sum x_i^2 - n(\bar{x})^2}}}}{\sqrt{\frac{(n-2)S^2}{\sigma^2}/(n-2)}}$$

er t-fordelt med $n-2$ frihetsgrader siden $(n-2)S^2/\sigma^2$ er χ^2_{n-2} fordelt og uavhengig av $\bar{Y}$ og $\hat{\beta}_1$ og dermed $\hat{\beta}_0$.

Under $H_0 : \beta_0 = 0$ er derfor $\frac{\sqrt{n}\hat{\beta}_0}{S \sqrt{\frac{\sum x_i^2}{\sum x_i^2 - n(\bar{x})^2}}}$ t-fordelt med $n-2$ frihetsgrader, og testen med forkastningsområde $|\frac{\sqrt{n}\hat{\beta}_0}{S \sqrt{\frac{\sum x_i^2}{\sum x_i^2 - n(\bar{x})^2}}}| > t_{1-\alpha/2, n-2}$ får nivå α.

Observerer

$$\frac{\sqrt{n}\hat{\beta}_0}{S \sqrt{\frac{\sum x_i^2}{\sum x_i^2 - n(\bar{x})^2}}} = \frac{\sqrt{17} \times 0.725}{\sqrt{45.56/15} \sqrt{\frac{16628.7}{16628.7 - (511.5)^2/17}}} = 0.47.$$

P-verdi $P(|t_{15}| > 0.47)$. Fra tabell $P(|t_{15}| > 0.258) = 0.8$ and $P(|t_{15}| > 0.536) = 0.6$, slik at p-verdien ligger mellom 0.6 og 0.8. Dette passer godt med resulatet fra punkt a) og b). Det er ikke så stor forskjell på de to tilpassede regresjonslinjene. Derfor betyr det ikke så mye å pålegge restriksjonen $\beta_0 = 0$. Testen $H_0 : \beta_0 = 0$ gir derfor heller ikke forkasting.

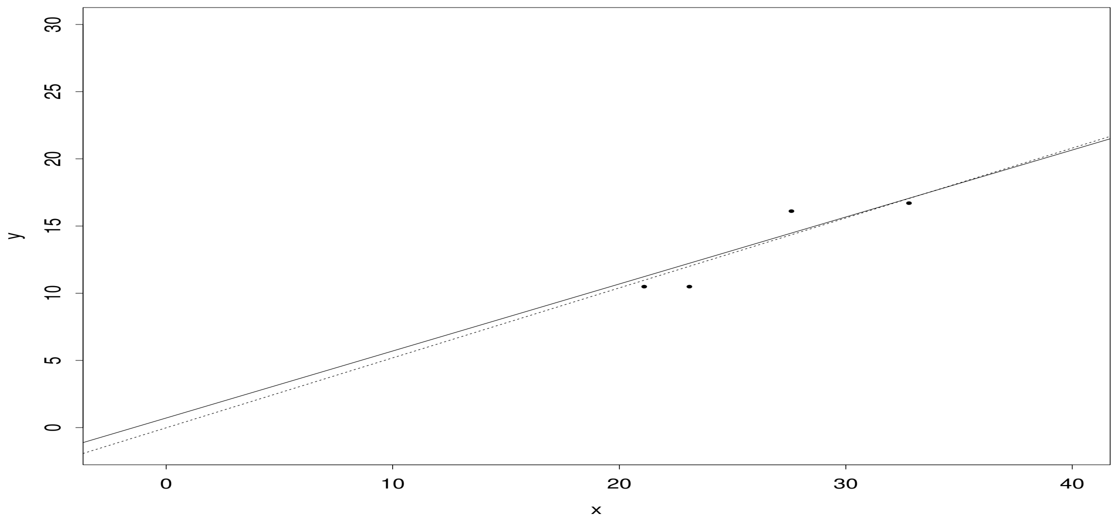

Figure 1: Regresjonslinjer.

Oppgave 4

Kvantiler i normalfordeling $N(0,1)$: $\Phi(x_p) = p, x_p = \Phi^{-1}(p)$
Kvantiler i logistisk fordeling: $p = 1/(1+e^{-y_p}), y_p = \log(p/(1-p))$.
Dette gir tabellen

p	$x_p = \Phi^{-1}(p)$	$y_p = \log(p/(1-p))$
0.1	$\Phi^{-1}(0.1) = -1.28$	$\log(0.1/0.9) = -2.20$
0.25	$\Phi^{-1}(0.25) = -0.67$	$\log(0.25/0.75) = -1.10$
0.5	$\Phi^{-1}(0.) = 0.0$	$\log(0.5(0.5) = 0$
0.75	$\Phi^{-1}(0.75) = 0.67$	$\log(0.75/0.25) = 1.10$
0.9	$\Phi^{-1}(0.9) = -1.28$	$\log(0.9/0.1) = -2.20$

Av plottet nedenfor ser man at kvantilene til den logistiske fordelingen har større absoluttverdi enn de tilsvarende kvantilene i standardnormalfordelingen. Den logistiske fordelingen har derfor "tyngre haler" enn normalfordelingen.

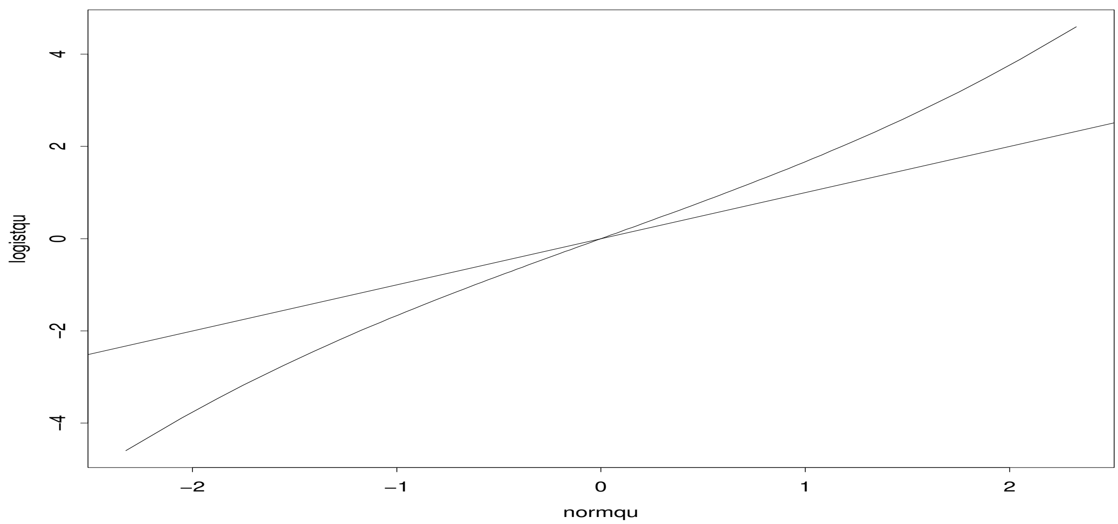

Figure 2: QQ-plott.