Oppgave 1

a) $$f(y|\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\beta x)^2\right\}$$

$$L(\beta|y_1 \dots y_n) = \prod_{i=1}^{n} f(y_i|\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2\right\}$$

$$= \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^n \exp\left\{-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2\right\} \propto$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2\right\}$$

b) Log-rimelighetsfunksjon

$$\ell(\beta) = \log L(\beta|y_1 \dots y_n) = n\log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] - \frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2$$

$$\frac{\partial}{\partial\beta}\ell(\beta) = 0 + \frac{1}{2\sigma^2}\sum(y_i-\beta x_i)\cdot 2x_i = 0$$

$$\Rightarrow \hat{\beta} = \frac{\sum Y_i x_i}{\sum x_i^2}$$

$\hat{\beta}$ er en lineær kombinasjon av Y_i som er normal-fordelte $\Rightarrow$ $\hat{\beta}$ også er normal fordelt

$$E(\hat{\beta}) = E\left[\frac{\sum Y_i x_i}{\sum x_i^2}\right] = \frac{\sum x_i E(Y_i)}{\sum x_i^2} = \beta \frac{\sum x_i^2}{\sum x_i^2} = \beta$$

$$Var(\hat{\beta}) = Var\left[\frac{\sum Y_i x_i}{\sum x_i^2}\right] = \frac{\sum x_i^2 Var(Y_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}$$

c) $H_0: \beta = b \quad H_1: \beta \neq b$ $\qquad Z = \dfrac{\hat{\beta} - b}{\sqrt{\sigma^2 / \sum X_i}} \sim N(0,1)$ under H_0

$$P(Z < -1.44) + P(Z > 1.44) = \alpha$$

$$P(Z < -1.44) = \alpha/2$$

$$P(Z < 1.44) = 0.075 \implies \underline{\alpha = 0.15}$$

d) $\lambda = \dfrac{L(b)}{L(\beta)}$

d) Verdier av $\frac{L(b)}{L(\hat{\beta})}$ nærme 1

antyder at observasjonene er kompatible med H_0, altså er observasjonene omtrent like godt forklart av b (H_0) som av $\hat{\beta}$ (SME).

Vi forkaster dermed H_0 dersom vi observerer $\lambda \leq \lambda^*$, der λ^* er kritisk verdi;

$$P(\lambda \leq \lambda^* \mid H_0) = \alpha$$

I vårt eksempel har vi

$$\lambda = \frac{L(b)}{L(\hat{\beta})} = \frac{e^{-\frac{1}{2\sigma^2}\sum_i (y_i - bx_i)^2}}{e^{-\frac{1}{2\sigma^2}\sum_i (y_i - \hat{\beta}x_i)^2}}$$

$$= e^{-\frac{1}{2\sigma^2}\sum_i [(y_i - bx_i)^2 - (y_i - \hat{\beta}x_i)^2]}$$

$$= e^{-\frac{1}{2\sigma^2}[\sum y_i^2 - 2b\sum y_i x_i + b^2 \sum x_i^2 - \sum y_i^2 + 2\hat{\beta}\sum y_i x_i - \hat{\beta}^2 \sum x_i^2]}$$

$$= e^{-\frac{1}{2\sigma^2}(\hat{\beta} - b)^2 \sum x_i^2}$$

$\lambda < \lambda^*$ $\quad (0 < \lambda^* < 1)$

$\downarrow$

$-\frac{1}{2\sigma^2}(\hat{\beta} - b)^2 \sum x_i^2 < \ln(\lambda^*)$ $\quad \ln(\lambda^*) = -c, \; c > 0$

$\downarrow$

$$\frac{1}{\sigma^2}(\hat{\beta}-b)^2\sum x_i^2 > 2c \qquad d = 2c$$

$\downarrow$

$$\frac{(\hat{\beta}-b)^2}{\sigma^2/\sum x_i^2} > d \qquad d > 0$$

Def: $Z = \dfrac{\hat{\beta}-b}{\sqrt{\sigma^2/\sum x_i^2}}$ og $k = \sqrt{d}$ $\quad Z \sim N(0,1)$ under H_0.

Fra sannsynlighetskvoteprinsippet skal vi altså forkaste dersom

$$\frac{\hat{\beta}-b}{\sqrt{\sigma^2/\sum x_i^2}} < -k \quad \text{eller} \quad \frac{\hat{\beta}-b}{\sqrt{\sigma^2/\sum x_i^2}} > k$$

med k definert slik at

$$P\left(-k \le \frac{\hat{\beta}-b}{\sqrt{\sigma^2/\sum x_i}} \le k \mid H_0\right) = 1-\alpha.$$

Ved $\alpha = 0.15$ er $k = z_{0.075} = 1.44$.

Oppgave 2

a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ $\quad \varepsilon_i \sim N(0, \sigma^2)$

$\varepsilon_i \perp \varepsilon_j \quad \forall i \neq j$

Antagelser: • lineær sammenheng x og y (el. E(Y)) ← ser vi i figur 1, med negativt stigningstall.
• normalfordelte støyledd med konstant varians (samme varians)
• uavhengige observervasjoner ← gitt i tekst, kan ikke se i figurer.

Residualer: * Kan være noe høyere varians for x = 1, men små forskjeller
* residualene ser ut til å være normalfordelte.

b) $\hat{\beta}_0 = 5.059 \quad \hat{\beta}_1 = -0.7959$

$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 = 5.059 + 0.7959 \cdot (1) = 4.2631$

$E(\hat{y}_0) = y_0 = \beta_0 + \beta_1 x_0$

$E(\hat{Y}_0 - Y_0) = 0$

$$Var(\hat{Y}_0 - Y_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$$

Vi har $\bar{x} = 0 \cdot 15 + 1 \cdot 20 / 35 = \frac{20}{35}$

Dermed er $\sum (x_i - \bar{x})^2 = 15\left(0 - \frac{20}{35}\right)^2 + 20\left(1 - \frac{20}{35}\right)^2$

Vi har også $s = 0.5813$.

95% prediksjonsintervall: $\hat{y}_0 \pm t_{0.025,33}\, s \sqrt{1 + \frac{1}{35} + \frac{(1-\bar{x})^2}{\sum (x_i - \bar{x})^2}}$

blir, med $t_{0.025,33} \approx 2$,

$[3.07, 5.45]$

3 ANOVA

a)

Source	df	SS	MS	F
Treat.	2	41.8	20.9	3.52
Error	27	160.52	5.94	
Total	29	202.34		

- $MSE = \frac{SSE}{dfE} = \frac{160.52}{27} = 5.94$

- $F = \frac{MSTR}{MSE} \Rightarrow MSTR = F \cdot MSE = 3.52 \cdot 5.94 = 20.9$

- $SSTR = MSTR \cdot df_{TR} = 20.9 \times 2 = 41.8$

- $SSTOT = SSE + SSTR = 202.34$

$H_0: \mu_1 = \mu_2 = \mu_3$ vs H_1: Ikke slik

$\alpha = 0.05$ $F_{0.95,2,27} = 3.35$ $\Rightarrow$ Forkast H_0

b) $H_0: \frac{\mu_1 + \mu_2}{2} \geq \mu_3$ vs $H_1: \frac{\mu_1 + \mu_2}{2} > \mu_3$

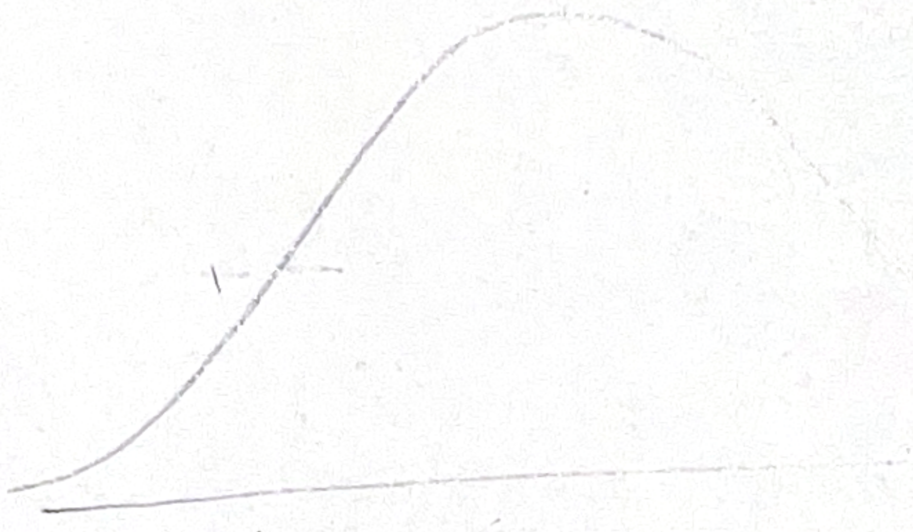
$C = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3$

$\hat{C} = \frac{1}{2} 15.38 + \frac{1}{2} 12.82 - 12.93 = 1.17$

$$SSC = \frac{\hat{C}^2}{\sum_{i=1}^{3} \frac{c_i^2}{n_i}} = \frac{1.17^2}{\frac{1}{4}\frac{1}{10} + \frac{1}{4}\cdot\frac{1}{10} + \frac{1}{10}} = \frac{1.17^2}{\frac{6}{40}} = 9.126$$

$$F = \frac{SSC/1}{SSE/27} = \frac{9.16}{5.94} = 1.54$$

$$T = \sqrt{F} = \sqrt{1.54} = 1.24$$



$$p\_verdi = P(T_{27} > 1.24) = 0.112$$

Ikke forkast

* kontrast brukes når man ~~bare~~ en interessert i et bestemt sammenligning.

Oppg. 4

a) $Y_i^{FØR} \sim f_i^{FØR}(y) \quad i=1..n$

$Y_i^{ETTER} \sim f_i^{ETTER}(y) \quad i=1..n$

Fortegntest

Antagelser

- Alle fordelinger $f_i^{FØR}(\cdot)$ og $f_i^{ETTER}(\cdot)$ er kontinuerlige (ikke nødvendigvis lik)
- $\text{median}(Y_i^{FØR}) = \tilde{\mu}_{FØR} \quad i=1...n$

  $\text{median}(Y_i^{ETTER}) = \tilde{\mu}_{etter} \quad i=1...n$

$H_0: \tilde{\mu}^{FØR} = \tilde{\mu}^{ETTER}$ vs $H_0: \tilde{\mu}^{FØR} > \tilde{\mu}^{ETTER}$

eller $p = P(Y_i^{FØR} > Y_i^{ETTER}) \quad i=1...n$

$H_0: p = \frac{1}{2}$ vs $H_1: p > \frac{1}{2}$

Wilcoxon test

Antagelser

- Alle fordelinger $f_i^{FØR}(\cdot)$ og $f_i^{ETTER}(\cdot)$ er kontinuerlige og symmetriske (ikke nødvendigvis like)
- $E(Y_i^{FØR}) = \mu^{FØR} \quad i=1...n$
- $E(Y_i^{ETTER}) = \mu^{ETTER} \quad i=1...n$

$H_0: \mu^{FØR} = \mu^{ETTER}$ vs $H_1: \mu^{FØR} > \mu^{ETTER}$

eller $\mu_D = E(Y_i^{FØR} - Y_i^{ETTER}) \quad i=1...n$

$H_0: \mu_D = 0$ vs $H_1: \mu_D > 0$

b) Fortegn test

U = antall par der $Y_i^{FØR} > Y_i^{ETTER}$

$U_{obs} = 9$

Under H_0 $U \sim Bin(10, 1/2)$

pvalue $= P(U \geq 9) \cong \cancel{0.0000}$ 0.01 med $\alpha = 0.05$ forkaster vi H_0

Wilcoxon test

	FØR	ETTER	Z	|Før − ETTER|	r
1	150	145	1	5	10
2	160	158	1	2	5
3	155	152	1	3	8,5
4	148	150	0	2	5
5	162	160	1	2	5
6	158	156	1	2	5
7	151	150	1	1	1,5
8	149	147	1	2	5
9	157	154	1	3	8,5
10	153	152	1	1	1,5

$$Z_i = \begin{cases} 1 & \text{hvis } Y_i^{FØR} > Y_i^{ETTER} \\ 0 & \text{ellers} \end{cases}$$

$r_i = \text{rank } |Y_i^{FØR} - Y_i^{ETTER}|$

$W = \sum r_i Z_i = 50$

$P(W \geq 50) = 1 - P(W \leq 49)$

$= 1 - 0.99 = 0.01$.

tabell

Forkaster H_0.