LØSNING.

Oppgave 1

a) $H_0: \mu_1 = \mu_2$ mot $H_1: \mu_1 \neq \mu_2$

$$T = \frac{\overline{Y}_{1.} - \overline{Y}_{2.}}{S_{12}\sqrt{\frac{1}{18}+\frac{1}{18}}}.$$ Forkast H_0 hvis $|T| \geq t_{0.025,34}$

$$S_{12}^2 = \frac{1743 + 4022.2}{34} = 13.0217^2$$

$$T = \frac{85.78 - 92.39}{13.0217 \cdot \sqrt{\frac{1}{18}+\frac{1}{18}}} = -1.5228$$

$$t_{0.025,34} \approx 2.032$$ dvs. Ikke forkast.

Ikke i tabell

b) Begrunnelse: "Muligens ikke normalfordelte obs."

$W_1 = 296.$ $\quad U_1 = 296 - \underbrace{\frac{18 \cdot 19}{2}}_{\text{minste mulige verdi for } W_1} = 125$

$U_2 = 18 \cdot 18 - 125 = 199$

Dermed: p-verdi $= 2 \cdot P(U_1 \leq 125)$
som beregnes ved tilnærming til normalford.:

②

$$E(U_1) = \frac{18 \cdot 18}{2} = 162 \quad , \quad Var(U_1) = \frac{18 \cdot 18 \cdot 37}{12} = 999$$

$$\Rightarrow \quad P(U_1 \leq 125) = \Phi\left(\frac{125 + 0.5 - 162}{\sqrt{999}} \right)$$

$$= \Phi(-1.1548) = 0.124$$

dvs P-value $= \underline{0.248} \geq 0.05$ dvs $\underline{\text{ikke}}$ forkast H_0.

(c) $$SSA = n \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= 18 \cdot \left[(85.78 - 91.76)^2 + (92.39 - 91.76)^2 + (97.11 - 91.76)^2 \right]$$

$$= 1166$$

$$SSE = 1743 + 4022.2 + 3103.8 = 8869$$

$$F = \frac{\frac{SSA}{2}}{\frac{SSE}{51}} = \frac{\frac{1166}{2}}{\frac{8869}{51}} = 3.3525$$

Skal forkaste H_0 hvis $F \geq f_{0.05, 2, 51} = 3.18$

ikke i tabell

dvs H_0 $\underline{\text{forkastes}}$!

d) Bonferroni:

$\mu_i - \mu_j$ er signifikant forskjellig fra 0 hvis

$$|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| \geq t_{51,\frac{0.05}{6}} \cdot s \sqrt{\frac{1}{18} + \frac{1}{18}}$$

6 pga. 3 hypoteser, og $\alpha/2 = 0.05/2$

Merk: Denne er $\frac{SSE}{51}$

$$= \underbrace{2.476}_{\text{ikke i tabell}} \cdot \sqrt{\frac{8869}{51}} \cdot \sqrt{\frac{2}{18}} = 10.88.$$

Dermed:

$\mu_1 - \mu_3 \neq 0$ (signifikant), siden

$$|\bar{y}_{1\cdot} - \bar{y}_{3\cdot}| > 10.88.$$

Ingen av de to andre forskjellene gir signifikans.

Tukeys metode: (Boka s. 420)

$|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}|$ må være større enn.

$$q(0.05, 3, 51) \cdot s \cdot \sqrt{\frac{1}{18}}$$

$$= 3.42 \cdot \sqrt{\frac{8869}{51}} \cdot \sqrt{\frac{1}{18}} = 10.63$$

App A22

Samme konklusjon!

Oppgave 2

(a) Modell og antagelser:

- $Y_i \sim N(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \sigma^2)$
- $Y_1, \dots, Y_{20}$ uavhengige.

Signifikant forklaringsgrad?

Tester $H_0: \beta_1 = \beta_2 = 0$ mot H_1: minst én $\neq 0$
dvs H_0: ingen regr. modell.

Dette testes ved $F = \dfrac{\frac{SSR}{2}}{\frac{SSE}{17}} = 10.63$
med p-verdi 0.01

Andel av variasjon forklart ved regresjonen:

$$R^2 = \frac{SSR}{SST} = \frac{121.483}{218.652} = \cancel{88.2}\ \underline{\underline{0.556}}$$

(b) Residualen er

$e_i = y_i - \hat{y}_i$

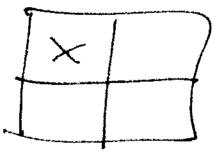
fra estimert regresjonsfunksjon "fitted values"

$E(e_i) = 0$
$Var(e_i) = (1 - h_{ii})\sigma^2$

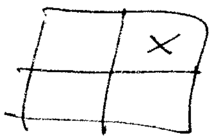
Forventes å oppføre seg som "uavhengige og $N(0,\sigma^2)$"
Dette utnyttes i plott.

- Normalplott
- Plott av $(\hat{y}_i, e_i)$
- Plott av (x_{ji}, e_i)
- Plott av (i, e_i)

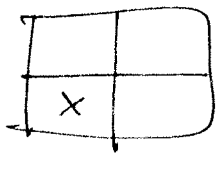
Figur 1:



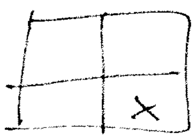
Normalplottet viser tendens til å ikke ligge på en rett linje, men heller som
Ideelt: Rett linje



U-form viser at modellen tilpasser ~~for høye verdi~~ lave verdier ~~for~~ ved lave responser, f.eks.
Ideelt: Punkter jevnt spredt på hver side av 0-aksen.



U-form: Gal modell for betydningen av x_1.
Ideelt: Som forrige plott.



Virker mer likt et ideelt plott.

Totalt sett: Modellen gir ingen god beskrivelse av dataene.

(c) Velger modellen med x_1, x_2, $x_1 * x_2$, $x_1 * x_1$.

Regr. analysen: Disse er alle signifikante, med p-verdier ≤ 0.008, mens $x_2 * x_2$ har høy p-verdi.

Best subsets: Modellen har høyest R^2_{adj}, lavest s, lavest C-p.

Forklart variasjon: $R^2 = 89.6\%$ mot 55.6% i (a)

Mallows C-p for (a)-modellen:

$$3 + \frac{5.716 - 1.6}{1.6}(20-3) = \underline{\underline{46.7}} \quad \text{(meget høy!)}$$

Matematisk grunn for ny modell: ~~Butter~~ 2. ordens Taylor-polynom for en funksjon $f(x_1, x_2)$ er av formen

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

(d) $H_0: \beta_{11} = \beta_{22} = 0$ mot H_1: ikke slik.

Testobs.

$$F = \frac{\frac{SSR - R(\beta_1, \beta_2, \beta_{12})}{2}}{\frac{SSE}{14}}$$

Fra SegSS i utskrift $\longrightarrow$ $= \dfrac{\frac{25.764 + 0.279}{2}}{1.6} = 8.14$

Med 5% sign.nivå forkastes hvis

$$F \geq 3.75 \quad (= f_{0.05,2,14})$$

dvs. Forkaster H_0!

---

y_0 : $x_1^0 = 80$, $x_2^0 = 10$

Har at

$$CI: \hat{y}_0 \pm t_{0.025,14} \cdot s\sqrt{x_0'(X'X)^{-1}x_0}$$

Det er oppgitt at $s\sqrt{} = 0.503$ så

$$x_0'(X'X)^{-1}x_0 = \left(\frac{0.503}{1.26511}\right)^2 = 0.1581$$

Dermed:

$$CI: 6.297 \pm 2.145 \cdot 0.503$$

dvs $(5.22, 7.38)$

mens

$$PI: 6.297 \pm 2.145 \cdot 1.26511 \cdot \sqrt{1 + 0.1581}$$

$(3.38, 9.22)$

Oppgave 3

Under H_0 er $X_i \sim \text{bin}(5, p)$

dvs $E(X_i) = 5p$

og $E\left(\sum_{i=1}^{280} X_i\right) = 280 \cdot 5p = 1400p$

Dermed er $\hat{p} = \dfrac{\sum_{i=1}^{280} X_i}{1400}$ en rimelig estimator

$$= \frac{69 + 2\cdot 35 + 3\cdot 17 + 4\cdot 1 + 5\cdot 1}{1400}$$

$$= \frac{199}{1400} = \underline{0.1421}$$

~~Da blir som~~

Har $P(X = x) = \binom{5}{x} p^x (1-p)^{5-x}$

Estimerte verdier er da:

$$\hat{p}_0 = (1-\hat{p})^5 = 0.4646$$
$$\hat{p}_1 = 5\hat{p}(1-\hat{p})^4 = 0.3849$$
$$\hat{p}_2 = 10\hat{p}^2(1-\hat{p})^3 = 0.1276$$
$$\hat{p}_3 = 10\hat{p}^3(1-\hat{p})^2 = 0.021$$
$$\hat{p}_4 = 5\hat{p}^4(1-\hat{p}) = 0.0018$$
$$\hat{p}_5 = \hat{p}^5 = 6 \cdot 10^{-5}$$

Forventede verdier blir dermed:

$$E_0 = 280 \cdot 0.4646 = 130.1$$
$$E_1 = 280 \cdot 0.3849 = 107.8$$
$$E_2 = 280 \cdot 0.1276 = 35.7$$
$$E_{\geq 3} = 280 - E_0 - E_1 - E_2 = 6.4$$

Dermed.

$$\chi^2 = \sum_{i=0}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(157-130.1)^2}{130.1} + \cdots$$

$$= 45,14$$

Under H_0 er $df = \underset{\text{ant celler}}{4} - \underset{\text{multinom.}}{1} - \underset{\text{estim. param}}{1} = 2$

des forkast på nivå 1% hvis $\chi^2 \geq$ ..

dvs klar forkastning