LØSNING TMA 4260 H04

Oppgave 1.

(a) Konstruerer ~~først~~ et fullstendig 2^3-design. Lar deretter $D = ABC$ være generatoren. Da blir definerende relasjon $\underline{I = D^2 = ABCD}$

	A	B	C	D	AB	Blokk
1	−	−	−	−	+	2
2	+	−	−	+	−	1
3	−	+	−	+	−	1
4	+	+	−	−	+	2
5	−	−	+	+	+	2
6	+	−	+	−	−	1
7	−	+	+	−	−	1
8	+	+	+	+	+	2

Alias-struktur:

$I = ABCD$

$\Downarrow$

$A = BCD$
$B = ACD$
$C = ABD$
$D = ABC$

$AB = CD$
$AC = BD$
$AD = BC$
$BC = AD$

Ser at A, B, C, D kan estimeres ukonfundert hvis alle trefaktorsamspill er null.

(b) Må velge én effekt som generator for blokkdeling. Ser at ABC ikke er mulighet, siden den er lik D (og det ~~krever at~~ gjør at D blir konfundert med blokkeffekten).

Velger et av tofaktor-samspillene, f.eks AB. (se figuren) som blokkgenerator.

Blokkene blir da

2	1
3	4
6	5
7	8

Ser at effekten CD blir konfundert med blokkeffekten. Forøvrig vil A, B, C, D bare være konfundert med tredje ordens samspill og derfor være estimerbare ukonfundert under antagelsen i (a).

Oppgave 2.

a) $H_0: \beta = 0$ mot $H_1: \beta < 0$.

$$T = \frac{\hat{\beta}}{\widehat{SD(\hat{\beta})}}$$

Her er $SD(\hat{\beta}) = \frac{\sigma}{\sqrt{S_{xx}}}$ og $\widehat{SD(\hat{\beta})} = \frac{s}{\sqrt{S_{xx}}}$

Under H_0 er $T \sim t_{27-2} = t_{25}$

dvs med 5% nivå forkastes H_0 hvis $T \leq -t_{25,\,0.05}$
$= -1.708$

Vi får $\underset{\substack{\uparrow \\ \text{MINITAB}}}{T =} -12.87$ som gir soleklar forkastning.

"Scatterplot" antyder at parti A følger en annen linje enn B og C. Dette ses også på residualplottene for 'Observation Order' (de 9 første residualene er alle negative), og spesielt på residualplottet mot "Parti".

b)

x_2	x_3	Parti
0	0	A
1	0	B
1	1	C

For A : $x_2 = x_3 = 0$: $\underline{y = 32.1 - 0.0601t}$

B : $x_2 = 1, x_3 = 0$: $y = 32.1 - 0.0601t + 3.97$
$\underline{= 36.1 - 0.0601t}$

C: $x_2 = x_3 = 1$: $y = 32.1 - 0.0601t + 3.97 - 0.51$

$\underline{y = 35.6 - 0{,}0601\,t}$

Parallelle regresjonslinjer fordi stign.tall
$\hat{\beta}_1 = -0.0601$ er felles.

Residualplott (basert på $e_i = y_i - \hat{y}_i$)

- Normalplottet ligger tilfredsstillende langs en rett linje. Histogrammet er også bra
- Residualer mot $\hat{y}_i$ og "order of data" viser ingen påfallende avvik fra "u.i.f. $N(0, \sigma^2)$"
  Nå er også "Residual Vs Parti" OK

Avvik som kan oppdages er f.eks.

- Normalplott ikke rettlinjet: ε_i-ene ikke normale.
- e_i mot $\hat{y}_i$ viser f.eks. økende/avtagende spredning når $\hat{y}_i$ vokser: Ikke-konstant varians σ^2 (f.eks $Var(\varepsilon_i)$ er en funksjon av prediktorer).
- e_i mot variabler (som her "Parti")
  Kan avsløre systematiske avvik ~~som gjør~~ fordi forventningsstruktur er galt representert.

c) $H_0: \beta_2 = \beta_3 = 0$ mot H_1: minst én er $\neq 0$

Testes ved

$$F = \frac{R(\beta_2, \beta_3 \mid \beta_1) / 2}{s^2}$$

$$= \frac{\frac{81.24 + 0.88}{2}}{1.605^2} = 15.94$$

Fra "Seq SS" i MINITAB

Under H_0 er $F \sim \text{Fisher}(2, 23)$

27-3-1

Med 5% nivå er kritisk verdi 3.42 så vi forkaster H_0 og konkluderer at regresjonslinjene ikke er sammenfallende.

d) $\hat{\beta}_3 \pm t_{23; 0.025} \cdot \widehat{SD}(\hat{\beta}_3)$

$-0.5078 \pm 2.069 \cdot 0.8681$

-0.5078 ± 1.7961

Intervall $(-2.3039, 1.2883)$

Siden $0 \in$ Intervall ville vi ikke forkastet $H_0: \beta_3 = 0$ mot $H_1: \beta_3 \neq 0$.

Dermed kan vi vurdere å ta ut ~~β_3~~ x_3 (som ville bety at partiene B og C beskrives ved samme regresjonslinje).

I utskriften er del en "Best Subsets Regression"
Her er R^2_{adj} størst for modellen der x_1 og x_2 er med. Denne har også lavest C_p, samt er sluttmodell i en Stepwise regression. Dette i tillegg til at $H_0: \beta_3 = 0$ aksepteres, taler for modellen med x_1 og x_2 inne.
(Skulle gjerne hatt residualplott også for denne, men det forventes at disse ville blitt temmelig like de vi har fra Modell 2).

Oppgave 3.

a) Svakheter: 3sigma-grensene er beregnet på normalfordelte eller iallefall symmetriske fordelinger. Kan også gi negative verdier for LCL.

Tabell gir at – hvis $V \sim \chi^2_3$ –

$$P(0.0243 < V < 16.266) = 0.998$$

Fra (3) har vi at $3S^2 \sim \chi^2_3$, dvs

$$P(0.0243 < 3S^2 < 16.266) = 0.998$$

som gir

$$P(0.090 < S < 2.329) = 0.998$$

Dermed
$$LCL = 0.090$$
$$UCL = 2.329$$

b) Nå er $\frac{3}{4}S^2 \sim \chi^2_3$

P(S faller utenfor)

$$= P(S < 0.090) + P(S > 2.329)$$

$$= P\left(\frac{3}{4}S^2 < \frac{3}{4}\cdot 0.090^2\right) + P\left(\frac{3}{4}S^2 > \frac{3}{4}\cdot 2.329^2\right)$$

$$= P(V < 0.0061) + P(V > 4.07)$$

$$= 0.000 + (1 - 0.746) = \underline{\underline{0.254}}$$

MINITAB EKSAKT

Forventet antall er $\dfrac{1}{0.254} = \underline{3.94}$

Sannsynl. fordl. for antallet, Y, er geometrisk,

$$P(Y=s) = 0.746^{s-1}\, 0.254 \quad ; \; s = 1, 2, \ldots$$

Oppgave 4.

Modell: $n = 2484$ deltagere klassifiseres med sannsynligheter (som har sum 1 over de 8 cellene)

	>Syk	Ikke	
Alder	p_{11}	p_{12}	$p_{1\cdot}$
i	p_{21}	p_{22}	$p_{2\cdot}$
	p_{31}	p_{32}	$p_{3\cdot}$
[illegible]	p_{41}	p_{44}	$p_{4\cdot}$
	$p_{\cdot 1}$	$p_{\cdot 2}$	1

H_0: $p_{ij} = p_{i\cdot}\, p_{\cdot j}$ for alle i, j.

Bruker uavhengighetstesten:

Chi-Square Test: C1; C2

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

           C1       C2  Total
    1      24     1355   1379
        61,07  1317,93
       22,499    1,043

    2      35      603    638
        28,25   609,75
        1,611    0,075

    3      21      192    213
         9,43   203,57
       14,186    0,657

    4      30      224    254
        11,25   242,75
       31,262    1,449

Total     110     2374   2484

Chi-Sq = 72,782; DF = 3; P-Value = 0,000
```

dvs H_0 forkastes på nivå 1%

~~P-verdi er [illegible]~~