NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET
INSTITUTT FOR MATEMATISKE FAG

Fagleg kontakt under eksamen:
Oddgeir Samset, telefon 99542779

EKSAMEN I EMNE SIF5068 INDUSTRIELL STATISTIKK

Måndag 3. desember 2001
Tid: 09.00 – 14.00

Hjelpemiddel: B - Alle trykte og handskrivne hjelpemiddel tillatne. Bestemt, enkel kalkulator tillate.

Sensuren fell 3. januar 2002

Oppgåve 1

Eit oljeselskap har utvikla eit nytt tilsetjingsstoff for motorolja si. Oljeselskapet har stor tru på at dette skal kunne føre til redusert bensinforbruk for bilar som nyttar denne motorolja.

For å undersøkje om dette held stikk blei følgjande eksperiment utført: Frå kvar av seks biltypar blei det valt ut to bilar kor den eine får fylt olje med tilsetjing, medan den andre får fylt olje utan tilsetjing. Kven av de to bilane frå kvar type som får olje med tilsetjing, blir avgjort med loddtrekning. Alle bilane kjører så ei løype som inkluderer både bykjøring og landevegskjøring, og gjennomsnittleg kjørelengde i mil pr. 10 liter bensin blir registrert. Data frå forsøket er gjeve i Tabell 1.1.

Tabell 1.1: Data frå forsøket med tilsetning i olje.

Biltype	Med	Uten	Diff
A	9.27	8.71	0.56
B	12.42	12.03	0.39
C	15.29	14.69	0.60
D	12.81	12.85	-0.04
E	10.41	10.42	-0.01
F	13.27	12.49	0.78

Det blir bestemt å gjennomføre ein test for å undersøkje om tilsetjingsstoffet aukar kjørelengda. Formuler dette som eit hypotesetestingsproblem, og vis korleis testen vert gjort viss vi kan gå ut frå at observasjonane

i) er uavhengige, med identisk, kontinuerleg og mogleg usymmetrisk men elles ukjent fordeling
ii) er uavhengige, med identisk, kontinuerleg og symmetrisk men elles ukjent fordeling

Kva blir konklusjonen på testen viss vi nyttar 10% nivå?
Kommenter moglege forskjellar mellom svara i i) og ii).

Kvifor er dette ein fornuftig forsøksplan? Kva er grunnen til at det blei nytta loddtrekning?

Oppgåve 2

Når sement størknar blir det avgjeve energi til omgjevnaden. For å undersøkje korleis mengde av fire kjemiske stoff x_1, x_2, x_3 og x_4 som inngår i sementen verker på avgjeve varme y frå sementen (kalorier pr. gram) blei det gjort 13 målingar som vist i Tabell 2.1.

Tabell 2.1: Observasjonar av avgjeve varme.

Obs.	y	x1	x2	x3	x4
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.3	11	66	9	12
13	109.4	10	68	8	12

Tabell 2.2: Korrelasjonsmatrisa.

	y	x1	x2	x3
x1	0.731			
x2	0.816	0.229		
x3	-0.535	-0.824	-0.139	
x4	-0.821	-0.245	-0.973	0.030

a)

I første omgang er vi interessert i å undersøkje korleis avgjeve varme y avheng av det kjemiske stoffet x_4. Utskrift frå regresjonsanalysen frå MINITAB er gjeve under i Tabell 2.3.

Set opp modellen og angje kva slags forutsetning som ligg til grunn for denne regresjonsanalysen.

Vi skal undersøkje om x_4 er viktig for y i denne modellen. Dette kan spesifiserast som eit hypotesetestingsproblem. Sett opp hypotesa. For denne spesielle modellen kan hypotesa testas på to ekvivalente måtar. Forklar korleis uttrykka for p-verdiane kjem fram for dei to måtane. Kva blir konklusjonen på testen dersom vi nyttar 1% nivå?

I utskrifta frå MINITAB har vi at $s = 8.964$ og $R\text{-}sq = 67.5\%$. Vis korleis s og $R\text{-}sq$ kjem fram og forklar kva dei er eit mål på.

Er det informasjon som du synes manglar i analysen av denne modellen? Forklar.

Tabell 2.3: Utskrift frå regresjonsanalysen frå MINITAB.

```
The regression equation is
y = 118 - 0.738 x4

Predictor        Coef    SE Coef       T      P
Constant      117.568      5.262   22.34  0.000
x4            -0.7382     0.1546   -4.77  0.001

S = 8.964      R-Sq = 67.5%     R-Sq(adj) = 64.5%

Analysis of Variance

Source           DF        SS        MS       F      P
Regression        1    1831.9    1831.9   22.80  0.001
Residual Error   11     883.9      80.4
Total            12    2715.8
```

b)

Vi skal nå prøve ein modell kor vi tek inn fleire av regresjonsvariablane.

Sjå nærare på korrelasjonsmatrisa i Tabell 2.2 over. Har denne eigenskapar som kan være til ulempe når fleire av regresjonsvariablane blir tekne med i modellen? Forklar.

For å undersøkje kva for variable som bør være med i modellen prøver vi med trinnvis regresjon. Forklar kva som skjer i trinn 1-4 i utskrifta i Tabell 2.4 under.

Tabell 2.4: Utskrift frå trinnvis regresjon frå MINITAB.

```
Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15
Response is y  on 4 predictors, with N = 13

    Step           1        2        3        4
Constant      117.57   103.10    71.65    52.58

x4            -0.738   -0.614   -0.237
T-Value        -4.77   -12.62    -1.37
P-Value        0.001    0.000    0.205

x1                       1.44     1.45     1.47
T-Value                 10.40    12.41    12.10
P-Value                 0.000    0.000    0.000

x2                               0.416    0.662
T-Value                           2.24    14.44
P-Value                          0.052    0.000

S               8.96     2.73     2.31     2.41
R-Sq           67.45    97.25    98.23    97.87
R-Sq(adj)      64.50    96.70    97.64    97.44
C-p            138.7      5.5      3.0      2.7
```

c)

Frå analysen i b) ser det ut som ein modell med både x_1 og x_2 gjev god tilpassing til dataene. Detaljer frå regresjonsanalysen for denne modellen frå MINITAB er gjeve under i Tabell 2.5.

Korleis vil du vurdera denne modellen opp mot modellen frå a)? Spesifiser dei kriteria som du legg til grunn for denne vurderinga og kommenter kvifor dei talar for eller mot modellane.

Dei sekvensielle kvadratsummane $SS(\beta_1 \mid \beta_0)$ og $SS(\beta_2 \mid \beta_0, \beta_1)$ er gjevne. Kva er $SS(\beta_1 \mid \beta_0, \beta_2)$?

Tabell 2.5: Utskrift frå regresjonsanalysen frå MINITAB.

```
The regression equation is
y = 52.6 + 1.47 x1 + 0.662 x2

Predictor        Coef     SE Coef        T       P
Constant       52.577       2.286    23.00   0.000
x1             1.4683      0.1213    12.10   0.000
x2            0.66225     0.04585    14.44   0.000

S = 2.406       R-Sq = 97.9%      R-Sq(adj) = 97.4%

Analysis of Variance
Source            DF         SS         MS        F       P
Regression         2     2657.9     1328.9   229.50   0.000
Residual Error    10       57.9        5.8
Total             12     2715.8

Source        DF      Seq SS
x1             1      1450.1
x2             1      1207.8
```

Oppgåve 3:

Det har blitt gjennomført eit laboratorie-forsøk for å sjå på effekten av 5 faktorar (A, B, C, D og E) på styrken av betong.

a)

Det forsøket som verkeleg blei gjennomført var ein halvfraksjon av eit 2^5 faktorielt design.

Tabell 3.1: 2^{5-1} fraksjonelt faktorielt design.

Obs.	A	B	C	D	E	y
1	-1	-1	-1	-1	-1	3.66
2	1	-1	-1	-1	1	2.74
3	-1	1	-1	-1	1	3.58
4	1	1	-1	-1	-1	3.90
5	-1	-1	1	-1	1	4.58
6	1	-1	1	-1	-1	3.82
7	-1	1	1	-1	-1	3.97
8	1	1	1	-1	1	5.08
9	-1	-1	-1	1	1	3.42
10	1	-1	-1	1	-1	2.75
11	-1	1	-1	1	-1	3.38
12	1	1	-1	1	1	4.26
13	-1	-1	1	1	-1	4.90
14	1	-1	1	1	1	4.12
15	-1	1	1	1	1	4.79
16	1	1	1	1	-1	5.61

Kva er generatoren og definerande relasjon for forsøket i Tabell 3.1?
Kva er resolusjonen til dette forsøket og kva vil den seie i praksis?
Skriv opp kva slags effektar som er konfundert med hovudeffekten C og med samspelet AB.

b)

Kvifor er randomisering svært viktig i gjennomføringa av eit forsøk? Korleis bør randomiseringa være gjort i dette tilfellet?

Gjev ei kort forklaring for bruken av normalplott til å identifisera dei signifikante (aktive) effektane.
Bruk utskrifta frå MINITAB i Tabell 3.2 og normalplottet av de estimerte effektane i Figur 3.3 til å finne dei signifikante (aktive) effektane i dette tilfellet. Gjev ei tolking av resultata frå analysen.

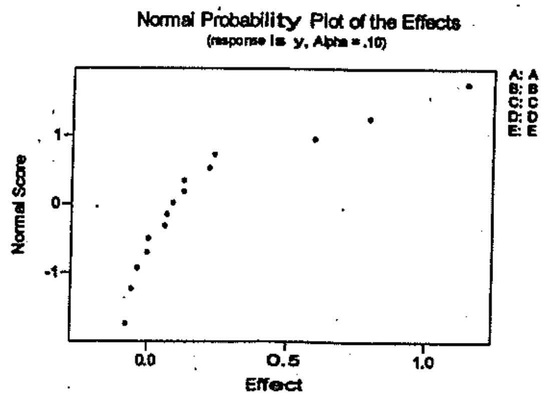
Tabell 3.2: Utskrift frå analysen av forsøket.

Estimated Effects and Coefficients for y

Term	Effect	Coef
Constant		4.035
A	-0.000	-0.000
B	0.573	0.286
C	1.148	0.574
D	0.238	0.119
E	0.073	0.036
A*B	0.783	0.391
A*C	0.098	0.049
A*D	0.063	0.031
A*E	-0.043	-0.021
B*C	-0.065	-0.033
B*D	0.140	0.070
B*E	0.140	0.070
C*D	0.255	0.128
C*E	-0.005	-0.003
D*E	-0.085	-0.043

Analysis of Variance for y (coded units)

Source	DF	Seq SS	Adj MS	F	P
Main Effects	5	6.825	1.3649	*	*
2-Way Interact	10	2.973	0.2973	*	*
Residual Error	0	0.000	0.0000		
Total	15	9.798			



Figur 3.3: Normalplott av estimerte effektar.

c)

I tillegg til dei 16 forsøka gjeve i Tabell 3.1 blei det samtidig gjennomført 3 enkeltforsøk i senter – det vil seie at alle nivåa er 0 for alle de 5 faktorane.

Sett opp regresjonsmodellen for dei 16+3 forsøka, og bruk dette til å forklare kvifor dei 3 enkelt-forsøka berre vil påverka estimatet av konstantleddet medan dei estimerte hoved- og samspelseffektane ikkje vil bli påverka.

Frå dei 3 enkeltforsøka fant ein at $SS_{ERROR} = 0.687$. Bruk dette til å lage eit 95% konfidensintervall for hovedeffekten C.

d)

På eit tidlegare tidspunkt i planlegginga av forsøket så anbefalte du å heller starte med ein kvartfraksjon av eit 2^5 faktorielt design med generatorar $D = ABC$ og $E = -AB$.

Skriv opp nivåkombinasjonane til alle enkeltforsøka for dette designet.
Kva blir definerende relasjon og kva er resolusjonen til designet?
Skriv opp alias-strukturen til designet under forutsetning om at trefaktor- og høgare ordens samspill kan neglisjeras.

I lys av analysen som blei gjennomført i b) kva for 'problemer' ville du ha kome opp i ved å nytta designet frå d)? Og korleis ville du ha løyst disse 'problema'?

e)

På eit endå tidlegare tidspunkt i planlegginga var det enkelte av deltakarne som heller ville gjennomføre eit 'ein-faktor-om-gongen' forsøk, dvs at ein og ein faktor blir variert om gangen medan alle dei andre faktorane blir heldt konstant.

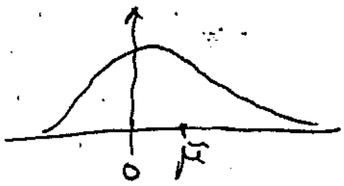
Skriv ned dei ulike punkta som du brukte for å overbevise dei andre deltakarne om at det var betre å starta med eit fraksjonelt faktorielt forsøk. Ta gjerne utgangspunkt i berre 3 faktorar for å forenkle problemstillinga, og bruk gjerne figurar i forklaringa.

Løsningsforslag

Oppgave 1:

Diff = Med − Uten   Diff ønskes stor

i) $H_0: \tilde{\mu} = 0$ mot $H_1: \tilde{\mu} > 0$ $\alpha = 10\%$



S − antall positive observasjoner av Diff

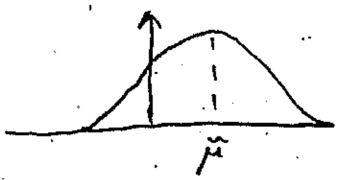
Under H_0: $S \sim bin(6, 1/2)$

Forkast H_0 dersom S blir "stor"

p-verdi $= P(S \geq 4 \mid H_0) = 1 - P(S \leq 3) = 1 - 0.656 = \underline{0.344}$

Data gir ikke grunnlag for å forkast H_0

ii) $H_0: \tilde{\mu} = 0$ mot $H_1: \tilde{\mu} > 0$ $\alpha = 10\%$



W_+ er antall positive ranger

Forkast H_0 dersom W_+ er "stor" eller W_- liten

−0.01	−0.04	0.39	0.56	0.60	0.78
1	2	3	4	5	6

$w_- = 3$ $w_+ = 18$

p-verdi $= P(W_- \leq 3 \mid H_0) = \underline{0.078}$

Data gir grunnlag for å forkaste H_0

Oppgave 2:

a) Modell: $Y_i = \beta_0 + \beta_1 X_{4i} + E_i$

Antagelser: E_i uavhengige $\forall i$
$E_i \sim N(0, \sigma^2)$

Ønsker å teste om X_4 har betydning for Y
$H_0: \beta_1 = 0$ mot $H_1: \beta_1 \neq 0$
Kan gjøres ved T-test eller F-test

- T-test:
  Estimator for β_1 er $\hat{\beta}_1$
  Har da at $(\hat{\beta}_1 - 0)/\sqrt{\widehat{Var}(\hat{\beta}_1)} \sim$ t-fordelt med n-2 fr.gr. under H_0

  Dette uttrykket er representert ved T-ratio i MINITAB
  p-value $= 2 \cdot P(T_{11} \leq -4.77 \mid H_0)$

- F-test:
  Har at $SS_T = SS_R + SS_E$
  Under H_0 vil $MS_R/MS_E \sim$ F-fordelt med 1 og 11 fr.gr.
  Dersom X_4 har betydning (H_1) forventer vi at MS_R blir "stor" i forhold til MS_E
  p-value $= P(F_{1,11} \geq 22.8 \mid H_0)$

p-verdi angir "sannsynligheten for å få det vi har observert eller noe mer ekstremt gitt at H_0 er riktig"
Hos oss er p-verdi ≈ 0 og indikerer derfor at det vi observerer er ekstremt i forhold til referansefordelingen under H_0
Dette indikerer at H_0 er feil – vi forkaster H_0.

$s^2 = MS_E = \frac{SS_E}{n-2}$ $\quad s = \sqrt{s^2}$

SS_E er delen av den totale variasjonen som modellen ikke fanger opp. MS_E er typisk et estimat av variansen σ^2 til den tilfeldige feilen.

$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$ R^2 beskriver andelen av den totale variasjonen som forklares ved regresjonsmodellen
$0 \leq R^2 \leq 1$ og høy verdi indikerer god tilpasning til dataene.

Kommentar:

- De to testene bygger på to forskjellige antagelser (om symmetri)
- Dersom vi kan anta symmetrisk fordeling som i ii) så kan vi lage en test med større teststyrke som lettere vil oppdage at medianen er større enn 0.
- Hvilken test som er riktig å bruke avhenger av den forhåndsinformasjon vi sitter inne med
- Bruk av en test som bygger på feil antagelse kan føre til at vi trekker feil konklusjon
  - ikke forkaster H_0 når H_1 er riktig
  - forkaster H_0 når H_0 er riktig

Hensiktsmessig forsøksplan:

- Forsøksplanen er en parplan
- Blokker er de ulike biltypene, og ved å bruke denne forsøksplanen trekker vi ut den variasjonen som skyldes forskjeller mellom de ulike biltypene
- Får testet ut mange biltyper.

Loddtrekning:

- Det vil alltid være en variasjon mellom biler av samme type - noen har høyere og noen lavere forbruk
- Loddtrekningen vil midle ut denne effekten slik at vi ikke får noe systematisk inn i forsøket som kan lure oss.

Informasjon som mangler er plott av dataene
- plott av y mot X_4
- plott av residualer mot $\hat{y}$ og X_4
- plott av residualer mot X_1, X_2 og X_3
- plott av residualer mot rekkefølge etc
- normalplott av residualer

} sjekk av antagelser

b) Typisk ved korrelasjonsmatrisen er
- y er sterkt korrelert med alle X-ene – mest med X_2 og X_4
- X_1 og X_3 sterkt korrelerte (-0.824)
- X_2 og X_4 sterkt korrelerte (-0.973)

Siden flere av X-ene er innbyrdes korrelerte kan vi få problemer med multikollinearitet. I så fall kan det resultere i at variansen til de estimerte parametrene $\hat{\beta}_i$ blir stor og vi får en dårlig modell.

Stepwise regresjon:
- Vi starter med en modell uten regr. variable
- Finn den regr. variable X_i som har størst $SS(\beta_i \mid \beta_0)$
  Dersom $SS(\beta_i \mid \beta_0)/MSE(\beta_i, \beta_0) > F_{enter}$ inkluderes X_i i modellen
- I hver iterasjon finn $SS(\beta_i \mid \beta_0, \dots \beta_{j-1})$ $i = j, \dots p$.
  for de variable som ikke er med i modellen
  Den X_i med størst $SS(\beta_i \mid \beta_0, \dots \beta_{j-1})$ inkluderes
  dersom $SS(\beta_i \mid \beta_0, \dots, \beta_{j-1})/MSE(\beta_0, \dots \beta_{j-1}\, \beta_i) > F_{enter}$

Hver gang en variabel inkluderes eller fjernes fra modellen beregner vi $SS(\beta_i \mid \beta_0, \dots \beta_j)$ for alle variablene som er med
Den X_i med minst $SS(\beta_i \mid \beta_0, \dots \beta_{j-1})$ fjernes fra modellen
dersom $SS(\beta_i \mid \beta_0, \dots \beta_j)/MSE(\beta_0, \dots \beta_j\, \beta_i) \leq F_{remove}$.

MINITAB:

Trinn 1 : X_4 tas med i modellen $T = -4.77$ $F = T^2 = 22.75$

Trinn 2 : X_1 — " — $T = 10.40$ $F = T^2 = 108.16$

Trinn 3 : Ingen fjerne fra modellen
X_2 tas med i modellen $T = 2.24$ $F = T^2 = 5.02$

Trinn 4 : X_4 fjernes fra modellen $T = -1.37$ $F = T^2 = 1.88$
Ingen flere variabler fjernes eller tas inn.

X_1 og X_2 forklarer sammen mer enn X_4 alene

c) Kriterium for valg av modell

Kriterium	Modell a)	Modell c)	
p-verdi for F-test om modell	0	0	
p-verdier for partielle T-tester	alle 0	alle 0	
s	8.96	2.41	favoriserer modell c)
R^2	67.45	97.87	
R^2_{adj}	64.50	97.44	
C_p	138.7	2.7	
Antall parametre i modell	1	2	
Plott for sjekk av antagelser	ikke tilgj.	ikke tilgj.	

$SS(\beta_1 \mid \beta_0) = 1450.1 \qquad SS(\beta_2 \mid \beta_0, \beta_1) = 1207.8$

Vet at partiell T-test for x_2

$T_1 = 14.44 \qquad F_1 = T_1^2 \qquad SS(\beta_2 \mid \beta_0, \beta_1) = F_1 \cdot MS_E \approx \underline{1207.3}$

Har da at

$SS(\beta_1 \mid \beta_0 \beta_2) = F_2 \cdot MS_E \approx \underline{\underline{847.7}}$

Oppgave 3

a) Generator: $E = -ABCD$
Definerende relasjon: $I = -ABCDE$

Resolusjon $R = V$
- hovedeffektene konfundert med 4-faktor samspill
- 2-faktor samspill konfundert med 3-faktor

$l_C \rightarrow C - ABDE$
$l_{AB} \rightarrow AB - CDE$

b) Randomisering er viktig for at den midler ut effekten av faktorer som vi ikke kan kontrollere eller ikke vet om. Vi unngår systematiske feil.

Alle 16 enkeltforsøkene bør gjøres i randomisert rekkefølge

Normalplott:

- normalfordeling
  $\hat{E} = \bar{Y}_+ - \bar{Y}_\div$ gjennomsnitt $\Rightarrow$ normalfordeling pga sentralgrense teoremet

- egenskaper
  $E[\hat{E}] = ?$
  $Var[\hat{E}] = \frac{4}{n}\sigma^2$ dvs lik for alle estimerte effekter

- Anta at alle effektene har forventning lik 0 normalplottet burde da bli en rett linje. Punkter som ikke ligger på denne ~~rette~~ linjen indikerer at de tilhørende effekt-estimat med forventning forskjellig fra 0., dvs aktive

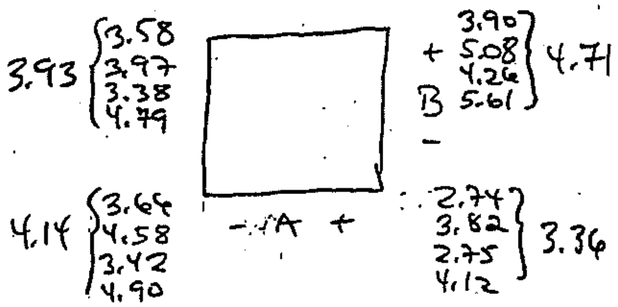
Det er 3 dominerende aktive effekter
C, B og AB

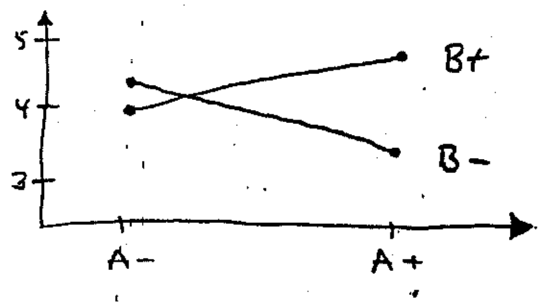
Tolkning:

- Faktor C inngår kun som hovedeffekt og kan tolkes direkte

  Å gå fra − til + nivå for faktor C gir en gjennomsnittlig økning i responsen på 1.15

- Faktorene A og B inngår som samspill og vi lager derfor samspillsplott





- Faktor B er spesielt viktig dersom A er på høyt nivå
- Sterkest betong : A+ B+ C+
- Svakest betong : A+ B− C−

c)

$$y = X\beta + E$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{16} \\ y_{17} \\ y_{18} \\ y_{19} \end{bmatrix} = \begin{bmatrix} & +/-1 & \\ 1 & 0 \cdots & 0 \\ 1 & 0 \cdots & 0 \\ 1 & 0 \cdots & 0 \end{bmatrix} \beta + E$$

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 1/19 & 0 & \cdots & 0 \\ 0 & 1/16 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & 1/16 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{19} y_i \\ \sum_{i=1}^{16} \pm y_i \\ \vdots \\ \sum_{i=1}^{16} \pm y_i \end{bmatrix}$$

dvs, kun konstantleddet blir påvirket!

$SS_{ERROR} = 0.687 \qquad DF = 3-1 = 2$

$\hat{\sigma}_Y^2 = \frac{SS_E}{2} = \underline{\frac{0.687}{2}}$

$\sigma_{eff}^2 = \frac{4}{16}\sigma_Y^2 = \underline{\frac{1}{4}\sigma_Y^2} \qquad \hat{\sigma}_{eff}^2 = 0.086$

$\hat{\sigma}_{eff} = 0.293$

95% konfidensintervall

$\hat{c} \pm t_{\frac{\alpha}{2},2} \cdot \hat{\sigma}_{eff}$

$1,148 \pm 4,30 \cdot \sqrt{\frac{1}{4}\frac{0.687}{2}}$

$1,148 \pm 1,26$

$\underline{[-0,112 \, , \, 2.408]}$

med dette estimatet av variansen blir faktisk ikke noen av effektene signifikante

Dette estimatet av σ_{eff}^2 stemmer dårlig overens med resten av analysen. Lager vi et estimat av σ_{eff}^2 basert på de 9 minste 2-faktorsamspillene får vi $\hat{\sigma}_{eff}^2 \approx 0.0146$. Man burde derfor sjekke opp om de 3 senterpunktene er riktig avlest og om SS_{ERROR} ble beregnet riktig.

d)

A	B	C	D	E
−	−	−	−	−
+	−	−	+	+
−	+	−	+	+
+	+	−	−	−
−	−	+	+	−
+	−	+	−	+
−	+	+	−	+
+	+	+	+	−

Generator: $D = ABC \quad E = -AB$

Definerende relasjon: $I = ABCD = -ABE = -CDE$

Resolusjon: $R = III$

Alias-struktur:

$l_A \rightarrow A - BE$ $\quad$ $l_{AC} \rightarrow AC + BD$
$l_B \rightarrow B - AE$ $\quad$ $l_{AD} \rightarrow AD + BC$
$l_C \rightarrow C - DE$
$l_D \rightarrow D - CE$
$l_E \rightarrow E - AB - CD$

Problemer: fra b) har vi at C, B og AB er aktive
Pga at vi nå har et resolusjon III design vil det være mange mulige tolkninger av resultatet

f.eks B, C, AB
E, AE, DE osv.

Det "naturlige" oppfølgingsdesignet ville nok vært et foldet design av resolusjon R=IV for å løse ut alle hovedeffektene fra 2-faktor-samspillene.
Dette ville skapt problemer her siden AB og CD da fortsatt ville vært konfundert.

Vi måtte valgt et design som kunne løst ut AB

dvs med generator $I = -ABCD = ABE = -CDE$

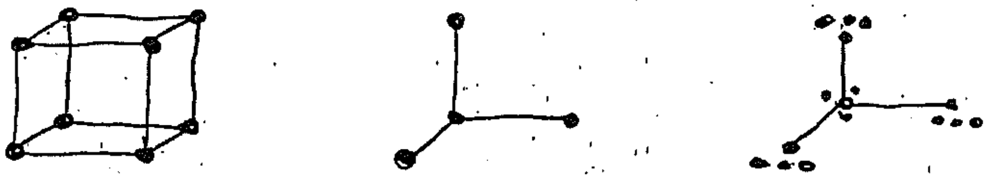
$l_E' \rightarrow E + AB - CD$ $\quad$ $(l_E' - l_E)/2 \rightarrow AB$
$l_B' \rightarrow B + AE$ $\quad$ $(l_B' + l_B)/2 \rightarrow B$
$l_C' \rightarrow C - DE$ $\quad$ denne får vi ikke løst opp

Vi står igjen med 2 alternative modeller

B, AB og C ← denne er nok mest realistisk
B, AB og DE

men trolig ville nok det anbefalte designet blitt en liten nøtt her...

e)



2^3-design   en-faktor

Problem med "en-faktor-om-gangen"

- får ikke estimert samspill
- må ha flere enkeltforsøk for å oppnå samme presisjon
- vanskeligheter med blokkdeling
- resultatene "gyldige" når de andre faktorene holdes konstant