Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Bo Lindqvist
Tlf. 975 89 418

BOKMÅL

EKSAMEN I FAG TMA4255
FORSØKSPLANLEGGING OG ANVENDTE STATISTISKE METODER

Fredag 5. desember 2008

Tid: 09:00–13:00

Tillatte hjelpemidler:
Alle trykte og håndskrevne hjelpemidler. Spesiell kalkulator.

Sensur: 5. januar 2009

Oppgave 1

Utbyttet av en kjemisk prosess ble studert i et pilotforsøk. Følgende faktorer ble betraktet:

Faktor		Faktornivåer	
		-1	1
A	Mengde av aktiv komponent	4 mol	5 mol
B	Surhet, pH	6	7
C	Reaksjonstid	2 timer	4 timer
D	Filtrering (første)	ingen	etter 1/2 time
E	Filtrering (andre)	ingen	etter 1 time

Det ble utført et 2^{5-2} fraksjonelt forsøk basert på et fullt 2^3-forsøk med faktorer A, B, C og med generatorer $D = AB$ og $E = AC$.

Responsen Y ble definert som utbytte målt relativt til et teoretisk maksimum. Responsene $Y_1, \ldots, Y_8$ i de 8 forsøkene antas uavhengige og normalfordelte med samme varians σ^2.

MINITAB-utskriften på neste side angir designet og de 8 responsene, samt alias-struktur og estimerte effekter. Du kan bruke denne utskriften i løsningen av oppgaven.

```
Data Display

Row  StdOrder   A   B   C   D   E      Y
  1         1  -1  -1  -1   1   1   69,3
  2         2   1  -1  -1  -1  -1   70,8
  3         3  -1   1  -1  -1   1   71,3
  4         4   1   1  -1   1  -1   73,2
  5         5  -1  -1   1   1  -1   77,5
  6         6   1  -1   1  -1   1   79,3
  7         7  -1   1   1  -1  -1   88,9
  8         8   1   1   1   1   1   91,2

Alias Structure

I + ABD + ACE + BCDE
A + BD + CE + ABCDE
B + AD + CDE + ABCE
C + AE + BDE + ABCD
D + AB + BCE + ACDE
E + AC + BCD + ABDE
BC + DE + ABE + ACD
BE + CD + ABC + ADE

Estimated Effects and Coefficients for Y (coded units)

Term       Effect     Coef
Constant           77,6875
A          1,8750   0,9375
B          6,9250   3,4625
C         13,0750   6,5375
D          0,2250   0,1125
E          0,1750   0,0875
B*C        4,7250   2,3625
B*E        0,0250   0,0125
```

a) Man er først interessert i å estimere hovedeffekten av faktor A. Forklar hvordan estimatet beregnes fra de gitte responsene. Gi en kommentar i lys av den oppgitte aliasstrukturen.

Vis også hvordan tofaktorsamspillet BC kan beregnes fra de gitte responsene og gi en kommentar. Hvilken generell fortolkning har et tofaktorsamspill?

Hva er de definerende relasjoner og hva er resolusjonen for designet i denne oppgaven? Hvilken praktisk fortolkning har resolusjonen?

Det ble i analysen antatt at faktorene D og E ikke har interaksjon med hverandre eller med noen av de andre faktorene i forsøket, dvs. at alle samspill der D eller E deltar kan settes til 0. Denne antagelsen skal gjøres i resten av oppgaven.

b) Hvilke hovedeffekter kan nå estimeres ukonfundert, og hvilke hovedeffekter er fremdeles konfundert med andre effekter?

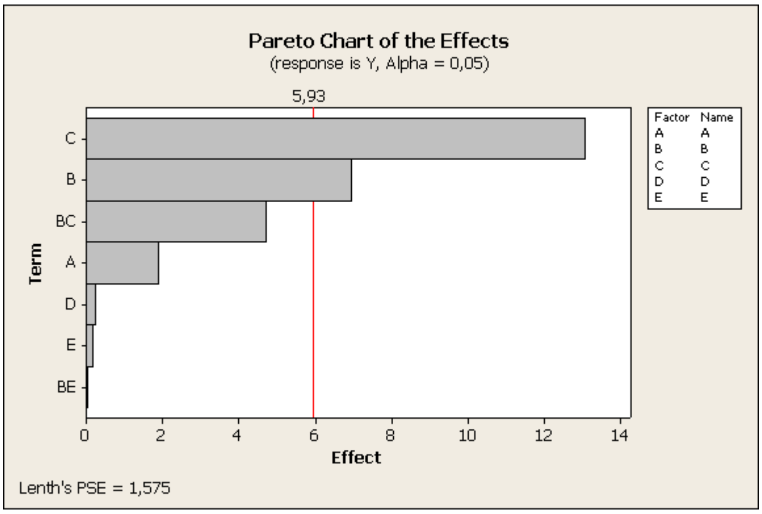
Finn variansen for en estimert effekt uttrykt ved σ^2.

Fra tidligere erfaring ble σ^2 antatt kjent, $\sigma^2 = 2$. Bruk dette til å avgjøre hvilke hovedeffekter og samspill som er signifikante når signifikansnivået velges til 5%.

Hva er det minste signifikansnivået som ville føre til at A har signifikant effekt?

c) I Pareto-diagrammet i figuren nedenfor brukes Lenth's PSE som estimat for standardavviket til de estimerte effektene.

Vis ved hjelp de estimerte effektene i MINITAB-utskriften hvordan man kommer fram til resultatet PSE = 1.575.



Oppgave 2

Et forskningslaboratorium skal undersøke vibrasjonen som oppstår når tynne plater av polyethylen utsettes for vind.

Til forsøket brukes plater av samme tykkelse og bredde, men med tre forskjellige lengder. Platene utsettes for fire forskjellige vindhastigheter, og antall svingninger i et bestemt tidsrom blir målt. For hver av de 12 kombinasjonene av lengde og vindhastighet foretas to forsøk.

Responsen for enkeltforsøkene betegnes Y_{ijk} og angir en hundredel av antall svingninger pr. sekund når plate nummer k med lengde x_{1i} (faktornivå A_i) blir utsatt for vind av hastighet x_{2j} (faktornivå B_j); for $i = 1, 2, 3;\ j = 1, 2, 3, 4$ og $k = 1, 2$.

Resultatet av forsøket er gitt i tabellen nedenfor. Enheten for lengde er *tommer*, mens enheten for vindhastighet er *fot/sekund*.

Lengde	Vindhastighet (x_{2j})			
(x_{1i})	$B_1 = 62.5$	$B_2 = 54.6$	$B_3 = 44.3$	$B_4 = 31.3$
$A_1 = 1.50$	50.5	46.0	36.5	23.0
	50.0	45.1	37.0	24.5
$A_2 = 1.75$	47.0	41.5	33.1	22.0
	48.0	42.0	34.1	24.2
$A_3 = 2.00$	45.5	39.4	30.8	20.3
	45.1	38.8	31.0	21.6

Man analyserer først dataene ved hjelp av to-faktor variansanalyse, der faktorene A og B har henholdsvis 3 og 4 nivåer. Du kan i punktene nedenfor bruke følgende utskrift fra MINITAB:

Two-way ANOVA: Y versus A; B

```
Source       DF       SS        MS         F
A             2   100,54    50,268     93,52
B             3  2145,40   715,134   1330,48
Interaction   6     8,99     1,498      2,79
Error        12     6,45     0,537
Total        23  2261,38
```

a) Sett opp modellen for observasjonene $\{Y_{ijk}\}$ som tar hensyn til et eventuelt samspill mellom lengdeeffekten (A) og vindhastighetseffekten (B).

Hvordan vil du teste om det er samspill mellom de to faktorene? Sett opp og begrunn en testobservator og finn kritisk verdi når signifikansnivået settes til 5%. Hva blir konklusjonen?

Er hovedeffektene av lengde og vindhastighet signifikante med 5% nivå? Begrunn svaret ved å vise til de beregnede testobservatorene i MINITAB-utskriften.

b) Sett opp uttrykkene for SSE og $SS(AB)$ i modellen fra punkt (a).

Vis ved hjelp av resultater fra pensum at SSE/σ^2 er kjikvadrat-fordelt med 12 frihetsgrader. (*Vink:* Bruk at hvis $V_1, \ldots, V_n$ er uavhengige observasjoner fra $N(\nu, \tau^2)$, så er $\sum_{k=1}^{n}(V_k - \bar{V})^2/\sigma^2$ kjikvadratfordelt med $n-1$ frihetsgrader).

Sett opp et uttrykk for estimatoren S^2 for σ^2 basert på SSE. Finn den numeriske verdi for S^2 i MINITAB-utskriften.

Anta nå at det ikke er noe samspill mellom faktorene A og B, dvs. at alle parametrene $(\alpha\beta)_{ij}$ i modellen i punkt (a) er lik 0. Det kan da vises at $SS(AB)/\sigma^2$ er kjikvadratfordelt med 6 frihetsgrader, og at $SS(AB)$ er uavhengig av SSE. (Du skal ikke gjøre dette.)

Vis hvordan en ny estimator for σ^2 kan utledes ved å bruke både SSE og $SS(AB)$. Finn estimatoren og regn ut det nye estimatet for σ^2 med de gitte dataene.

Oppgave 3

Situasjonen er den samme som i Oppgave 2. Det antas nå at

$$Y_{ijk} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2j} + \epsilon_{ijk} \qquad (1)$$

for $i = 1, 2, 3;\ j = 1, 2, 3, 4$ og $k = 1, 2$.

En utskrift fra en regresjonsanalyse med MINITAB basert på dataene i Oppgave 2 er gitt på neste side. Du kan bruke denne i løsningen av oppgaven.

a) Forklar kort hvorfor denne modellen kan behandles som en multippel lineær regresjonsmodell med 24 observasjoner og forklaringsvariable $x_1 =$ lengde og $x_2 =$ vindhastighet.

Hvilke antagelser gjøres i modellen (1)?

Hvor stor del av variasjonen i dataene er forklart ved hjelp av modellen? Har modellen signifikant forklaringsgrad?

Man vil teste hypotesene $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ og $H_0 : \beta_2 = 0$ mot $H_1 : \beta_2 \neq 0$ i denne modellen.

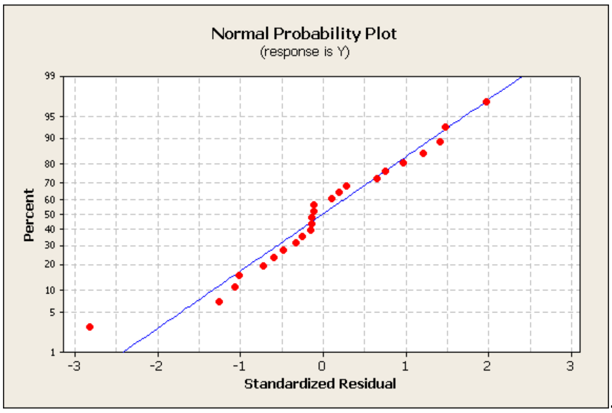
Sett opp de relevante testobservatorene og utfør de to testene med signifikansnivå 1%.

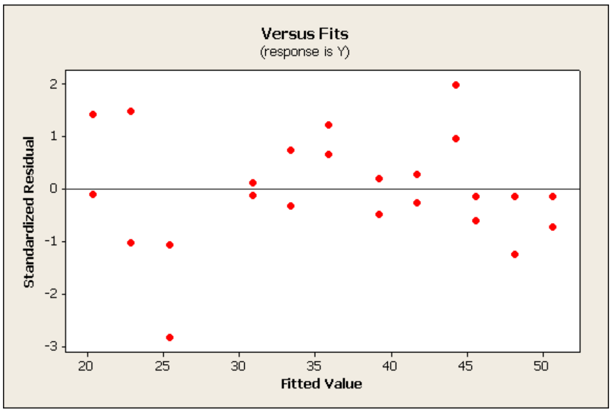
I hvilken grad støtter resultatene av denne testingen opp om konklusjonen fra variansanalysen i Oppgave 2?

Hva er likheter og forskjeller mellom modellen i denne oppgaven og modellen i Oppgave 2?

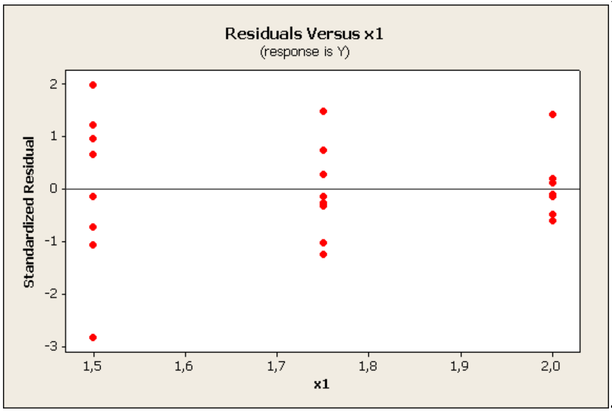
Regression Analysis: Y versus x1; x2

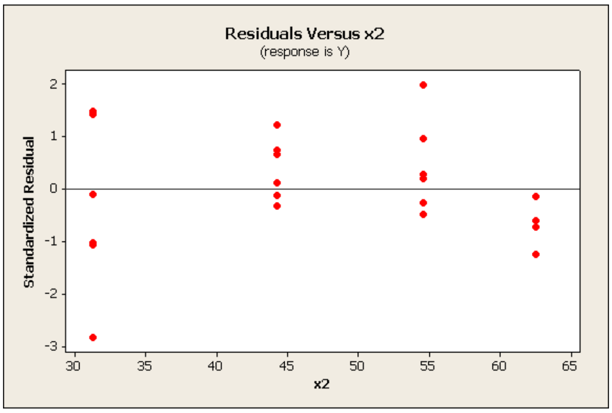
```
The regression equation is
Y = 15,1 - 10,0 x1 + 0,808 x2


Predictor        Coef   SE Coef       T
Constant       15,140     1,848    8,19
x1           -10,0250    0,9466  -10,59
x2            0,80842   0,01653   48,89

Analysis of Variance

Source          DF      SS      MS        F      P
Regression       2  2242,6  1121,3  1251,40  0,000
Residual Error  21    18,8     0,9
Total           23  2261,4
```





,





,

b) Fire ulike residualplott er gitt på forrige side.

Hva menes her med standardiserte ("studentized") residualer?

Forklar kort hva som plottes i hvert enkelt diagram, og gi en vurdering av resultatene med hensyn til tilpasning mellom data og modell.

Har du forslag til enkle utvidelser av modellen som muligens kan forbedre tilpasningen? Det kreves ingen nye analyser.

Oppgave 4

La X være antall millioner omdreininger til tretthetsfeil ved utprøving av en type kulelager. Det er testet 65 slike kulelager og resultatet er oppsummert i tabellen nedenfor.

Intervall	Antall observasjoner
$< 0, 40]$	7
$< 40, 60]$	14
$< 60, 80]$	18
$< 80, 100]$	15
$< 100, 120]$	8
$< 120, \infty]$	3

Man vil undersøke om sannsynlighetsfordelingen for X kan antas å være en Weibull-fordeling med kumulativ fordelingsfunksjon

$$F(x) = P(X \leq x) = 1 - e^{-\left(\frac{x}{80}\right)^2} \text{ for } x > 0. \tag{2}$$

a) Bruk tallmaterialet i tabellen til å utføre en test for nullhypotesen at X har fordelingen (2). La signifikansnivået være 5%. Hva blir konklusjonen?

Forklar kort hvordan du ville gått fram dersom nullhypotesen var at X er Weibull-fordelt, uten angivelse av verdier på parametre, og du hadde de samme dataene. Du skal ikke gjøre noen nye beregninger her.

(Weibull-fordelingen er generelt gitt ved en kumulativ fordelingsfunksjon

$$F(x; \theta, \alpha) = 1 - e^{-(\frac{x}{\theta})^\alpha} \text{ for } x > 0,$$

der α, θ er positive parametre.)