Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

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

EKSAMEN I FAG TMA4260 INDUSTRIELL STATISTIKK

Onsdag 1. desember 2004

Tid: 09:00–13:00

Tillatte hjelpemidler:
Alle trykte og håndskrevne hjelpemidler, samt kalkulatoren HP30S.

Sensur: 22. desember 2004

BOKMÅL

Oppgave 1

a) Konstruer et 2^{4-1}-design for de fire faktorene A, B, C og D når den definerende relasjon er I = ABCD.

Hva er henholdsvis generatoren og resolusjonen til designet?

Sett opp alias-strukturen og forklar kort hva denne uttrykker.

Under hvilke betingelser om samspillene kan vi estimere alle de fire hovedeffektene ukonfundert?

b) Anta at designet i punkt (a) skal deles i to like store blokker. Foreslå en generator for denne blokkdelingen slik at hovedeffektene fremdeles kan estimeres ukonfundert under betingelsene til slutt i punkt (a).

Sett opp de blokkene du får.

Hvilke(n) effekt(er) blir konfundert med blokkeffekten?

Oppgave 2

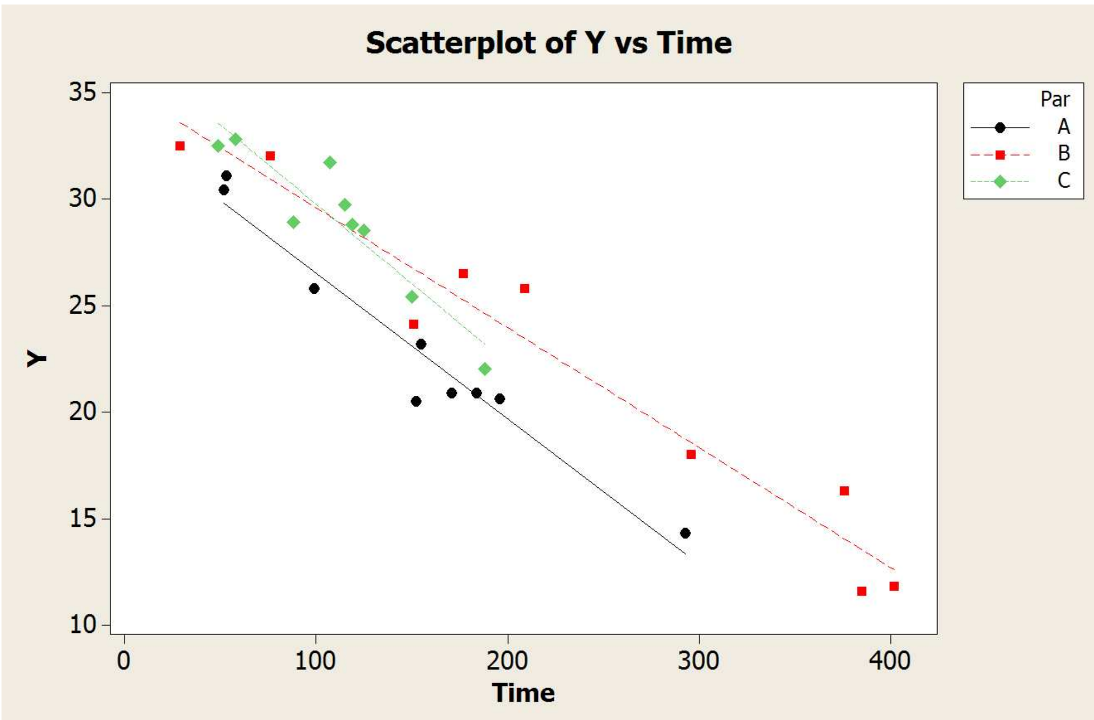
Et apparat er konstruert for å sende en jevn strøm av et hormon til en pasient uten at apparatet blir etterfylt. For $n = 27$ slike apparater har vi målinger av y=mengde hormon igjen i apparatet etter fjerning fra pasienten (i milligram), og t=den tid pasienten er blitt tilført hormonet (i timer). Apparatene kan ha forskjeller som skyldes at de kommer fra ulike leveransepartier. Dette er tatt hensyn til ved at man har valgt ut tilfeldig 9 apparater fra hvert av tre partier A,B og C.

Dataene er gitt i tabellen nedenfor, mens sammenhørende verdier av t (*Time*) og y er tegnet grafisk i figuren nederst på siden (med en rett linje estimert for hvert av partiene).

Parti	j	t_j	y_j
A	1	99	25.8
A	2	152	20.5
A	3	293	14.3
A	4	155	23.2
A	5	196	20.6
A	6	53	31.1
A	7	184	20.9
A	8	171	20.9
A	9	52	30.4

Parti	j	t_j	y_j
B	10	376	16.3
B	11	385	11.6
B	12	402	11.8
B	13	29	32.5
B	14	76	32.0
B	15	296	18.0
B	16	151	24.1
B	17	177	26.5
B	18	209	25.8

Parti	j	t_j	y_j
C	19	119	28.8
C	20	188	22.0
C	21	115	29.7
C	22	88	28.9
C	23	58	32.8
C	24	49	32.5
C	25	150	25.4
C	26	107	31.7
C	27	125	28.5



Man velger først å analysere dataene med en enkel lineær regresjonsmodell uten å ta hensyn til hvilket parti hver enkelt observasjon kommer fra. Modellen som brukes er da

$$y_i = \alpha + \beta t_i + \epsilon_i \tag{1}$$

for $i = 1, \ldots, 27$, der $\epsilon_1, \ldots, \epsilon_{27}$ er uavhengige og normalfordelte med forventning 0 og varians σ^2. Her er α, β og σ ukjente parametre.

Resultatet av en MINITAB-kjøring med denne modellen er gitt på sidene 6-7.

a) Man ønsker en test for

$$H_0: \ \beta = 0 \text{ mot } H_1: \ \beta < 0.$$

Sett opp testobservatoren for en t-test for denne situasjonen. For hvilke verdier av testobservatoren skal H_0 forkastes hvis signifikansnivået settes til 5%?

Bruk opplysningene fra MINITAB-kjøringen til å finne verdien på testobservatoren. Hva blir konklusjonen på testen?

Studer residualplottene for modellen (1), samt plottet på forrige side. Forklar hvorfor plottene kan tyde på at de tre partiene ikke bør beskrives av den samme modellen (1).

Man velger nå å utvide modellen til å tillate at de tre partiene har ulike regresjonslinjer. Siden de tre estimerte linjene i figuren på forrige side er nær parallelle, antar man at de tilsvarende teoretiske linjene er parallelle, men at de kan ha ulike skjæringspunkt med y-aksen.

Det innføres nå tre forklaringsvariabler (prediktorer) x_1, x_2, x_3. Her er x_1 den oppgitte tiden som pasienten er blitt tilført hormonet, mens x_2 og x_3 er variabler som definerer hvilket parti en observasjon kommer fra. Mer presist defineres

$$x_2 = \begin{cases} 1 & \text{hvis observasjonen kommer fra parti B eller C} \\ 0 & \text{ellers} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{hvis observasjonen kommer fra parti C} \\ 0 & \text{ellers} \end{cases}$$

Modellen for dataene kan nå skrives

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i \tag{2}$$

for $i = 1, \ldots, 27$, der $\epsilon_1, \ldots, \epsilon_{27}$ er uavhengige og normalfordelte med forventning 0 og varians σ^2, og $\beta_0, \beta_1, \beta_2, \beta_3, \sigma$ er ukjente parametre.

Utskriften fra en MINITAB-kjøring med denne modellen er gitt på sidene 8-11. Du skal bruke resultater fra denne i besvarelsen av resten av oppgaven.

b) Hvilke verdier av paret (x_2, x_3) svarer til henholdsvis parti A, B og C?

Sett opp ligningene for de estimerte regresjonslinjene svarende til hvert av partiene A, B og C.

Hvordan ivaretas antagelsen om parallelle regresjonslinjer for de tre partiene?

Gi en vurdering av modell (2) ved å studere residualplottene gitt på side 11. Forklar kort hvilke avvik fra modellen som generelt kan oppdages fra denne typen plott.

c) Man ønsker å teste nullhypotesen at de tre teoretiske regresjonslinjene svarende til partiene A, B og C er sammenfallende, mot alternativet at minst to av dem er forskjellige. (Antagelsen om at de er parallelle skal fortsatt gjelde).

Formuler nullhypotesen ved hjelp av β-parametrene i modellen. Sett opp en testobservator og beregn dens verdi.

Hva blir konklusjonen dersom du bruker signifikansnivå 5%?

d) Finn et 95 % konfidensintervall for β_3.

Det viser seg at dette intervallet inneholder verdien 0. Hvorfor gir dette en indikasjon på at variabelen x_3 bør vurderes tatt ut av modellen?

Hvilken modell vil du velge for dataene ut fra MINITAB-utskriften? Begrunn valget.

Oppgave 3

Ved normal produksjon av en bestemt sylindrisk motordel antas diameteren X å være normalfordelt med forventning μ (mm) og standardavvik σ (mm).

For å kontrollere produksjonen plotter man estimerte standardavvik S ved gitte tidspunkter i et såkalt S-diagram. Anta at det ved hvert slikt tidspunkt gjøres n uavhengige observasjoner $X_1, X_2, \ldots, X_n$ av X. Da er altså

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad \text{der } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Du kan bruke i denne oppgaven at

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \tag{3}$$

der χ^2_ν som vanlig betyr kjikvadratfordelingen med ν frihetsgrader.

Du får også bruk for en tabell over χ^2_3 (gitt på side 12 i oppgavesettet).

Kontrolldiagrammene som bedriften har brukt hittil, har spesifisert øvre og nedre kontrollgrenser gitt ved "3 sigma"-grensene $\mathrm{E}(S) \pm 3\sqrt{\mathrm{Var}(S)}$. Man ønsker nå isteden å bruke egenskapen (3) til å lage kontrollgrenser LCL og UCL for S-diagrammet. Disse skal være slik at ved normal produksjon er

$$P(\mathrm{LCL} < S < \mathrm{UCL}) = 1 - q \tag{4}$$

for en på forhånd oppgitt (liten) sannsynlighet q. Dessuten ønsker man at $P(S < \mathrm{LCL}) = P(S > \mathrm{UCL})$

a) Hvilke svakheter kan du se ved bruk av "3 sigma"-grensene i et S-diagram?

Finn numeriske verdier for LCL og UCL i (4) når $n = 4$, $q = 0.002$ og σ er kjent, $\sigma = 1.0$.

b) Anta at det virkelige standardavviket for X er dobbelt så stort som tidligere antatt, dvs. $\sigma = 2.0$ istedenfor 1.0.

Vis at det da er en sannsynlighet på ca. 0.25 for at en enkelt observasjon av S vil falle utenfor kontrollgrensene $(\mathrm{LCL}, \mathrm{UCL})$ fra punkt (a).

Hva er det forventede antall nye S som må observeres før den økte variasjonen avsløres av S-diagrammet? Hvilken sannsynlighetsfordeling har dette antallet?

Oppgave 4

British Medical Journal publiserte i 1985 resultatet av en epidemiologisk spørreundersøkelse der hensikten var å finne ut om snorking kan være en risikofaktor for hjertesykdom. Et antall på 2484 deltakere ble klassifisert etter ektefellenes rapportering om hvor mye de snorket, og etter om de hadde en hjertesykdom eller ikke.

Resultatet er gitt i følgende tabell:

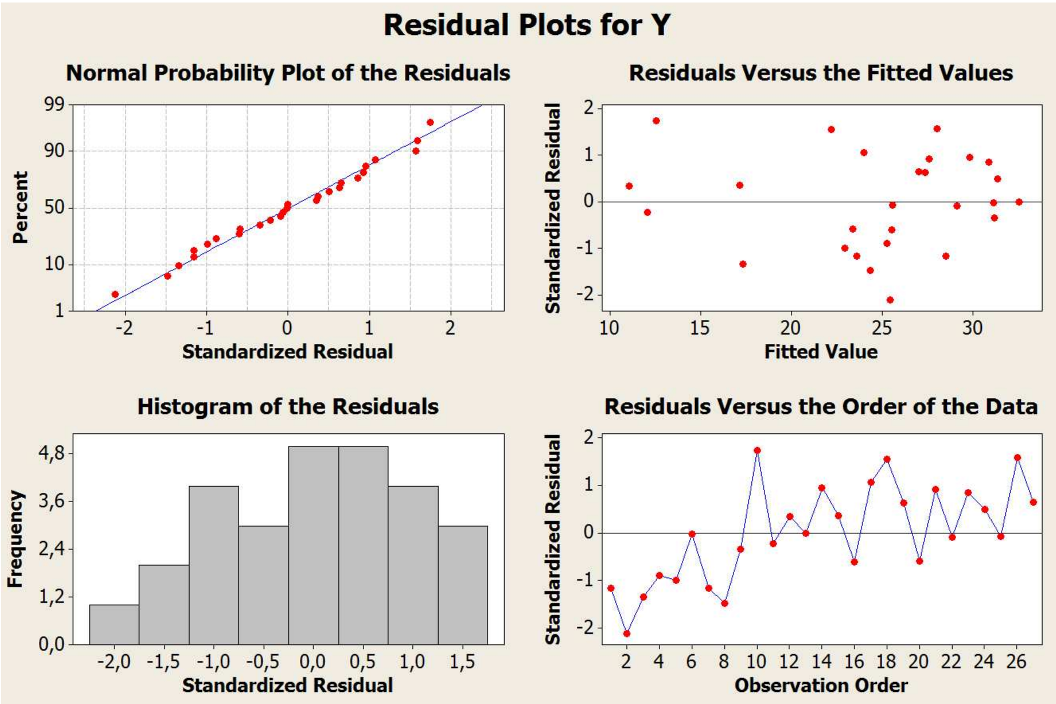
Snorking	Hjertesykdom	Ikke hjertesykdom
Aldri	24	1355
Av og til	35	603
Nesten hver natt	21	192
Hver natt	30	224

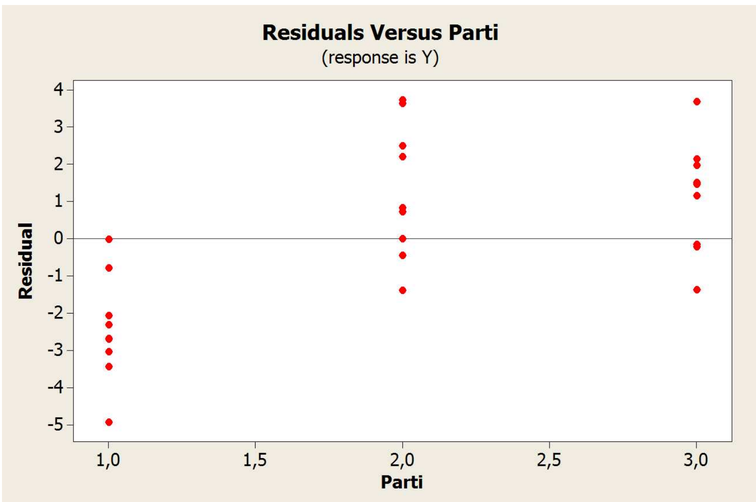
Tyder resultatet på at det er noen sammenheng mellom snorkemønster og hjertesykdom? Formuler selv modell og hypoteser og utfør en test. Hva blir konklusjonen dersom signifikansnivået settes til 1%?

MINITAB-utskrift for modell (1)

```
Regression Analysis: Y versus Time

The regression equation is
Y = 34,2 - 0,0574 Time


Predictor        Coef   SE Coef       T      P
Constant      34,1675    0,8672   39,40  0,000
Time        -0,057446  0,004464  -12,87  0,000


S = 2,37816   R-Sq = 86,9%   R-Sq(adj) = 86,4%


Analysis of Variance

Source          DF       SS      MS       F      P
Regression       1   936,54  936,54  165,59  0,000
Residual Error  25   141,39    5,66
Total           26  1077,93
```

Residualplott for modell (1)





Her svarer 1.0, 2.0 og 3.0 på x-aksen til partiene A, B og C henholdsvis.

MINITAB-utskrift for modell (2)

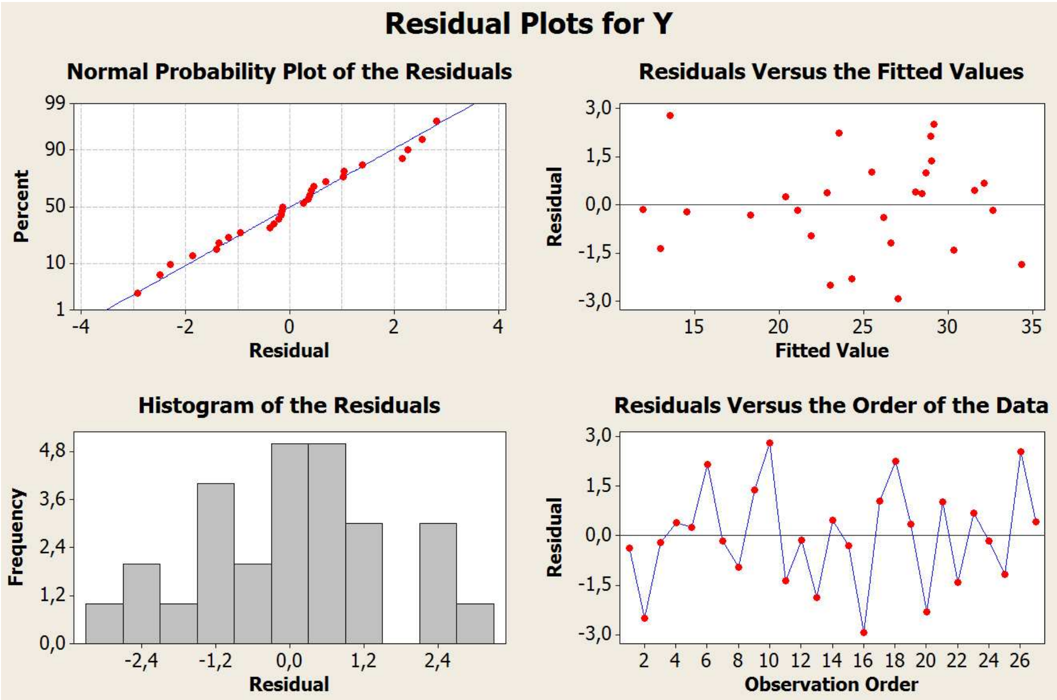
Data Display

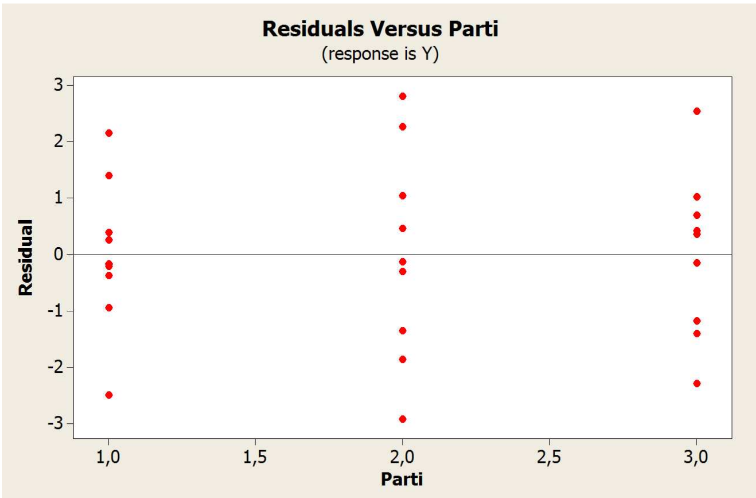
Row	Par	Y	x1	x2	x3
1	A	25,8	99	0	0
2	A	20,5	152	0	0
3	A	14,3	293	0	0
4	A	23,2	155	0	0
5	A	20,6	196	0	0
6	A	31,1	53	0	0
7	A	20,9	184	0	0
8	A	20,9	171	0	0
9	A	30,4	52	0	0
10	B	16,3	376	1	0
11	B	11,6	385	1	0
12	B	11,8	402	1	0
13	B	32,5	29	1	0
14	B	32,0	76	1	0
15	B	18,0	296	1	0
16	B	24,1	151	1	0
17	B	26,5	177	1	0
18	B	25,8	209	1	0
19	C	28,8	119	1	1
20	C	22,0	188	1	1
21	C	29,7	115	1	1
22	C	28,9	88	1	1
23	C	32,8	58	1	1
24	C	32,5	49	1	1
25	C	25,4	150	1	1
26	C	31,7	107	1	1
27	C	28,5	125	1	1

```
Regression Analysis: Y versus x1; x2; x3

The regression equation is
Y = 32,1 - 0,0601 x1 + 3,97 x2 - 0,508 x3


Predictor       Coef   SE Coef       T      P
Constant     32,1316    0,7483   42,94  0,000
x1         -0,060136  0,003474  -17,31  0,000
x2            3,9735    0,8097    4,91  0,000
x3           -0,5078    0,8681   -0,58  0,564


S = 1,60530   R-Sq = 94,5%   R-Sq(adj) = 93,8%


Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  1018,66  339,55  131,76  0,000
Residual Error  23    59,27    2,58
Total           26  1077,93


Source  DF  Seq SS
x1       1  936,54
x2       1   81,24
x3       1    0,88



Best Subsets Regression: Y versus x1; x2; x3

Response is Y

                       Mallows          x x x
Vars  R-Sq  R-Sq(adj)      C-p       S  1 2 3
   1  86,9       86,4     31,9  2,3782  X
   1  22,4       19,3    301,4  5,7827      X
   2  94,4       94,0      2,3  1,5831  X X
   2  88,7       87,8     26,1  2,2485  X   X
   3  94,5       93,8      4,0  1,6053  X X X
```

```
Stepwise Regression: Y versus x1; x2; x3

  Alpha-to-Enter: 0,15  Alpha-to-Remove: 0,15


Response is Y on 3 predictors, with N = 27


Step                1        2
Constant        34,17    31,98

x1            -0,0574  -0,0591
T-Value        -12,87   -19,80
P-Value         0,000    0,000

x2                        3,70
T-Value                   5,69
P-Value                  0,000

S                2,38     1,58
R-Sq            86,88    94,42
R-Sq(adj)       86,36    93,95
Mallows C-p      31,9      2,3
```

Residualplott for modell (2)





Her svarer 1.0, 2.0 og 3.0 på x-aksen til partiene A, B og C henholdsvis.

MINITAB-utskrift for Oppgave 3

```
Inverse Cumulative Distribution Function of Chi-Square with 3 DF
```

$P(X \leq x)$	x
0.0001	0.0052
0.0010	0.0243
0.0020	0.0387
0.0050	0.0717
0.0100	0.1148
0.0200	0.1848
0.1000	0.5844
0.1500	0.7978
0.2000	1.0052
0.2500	1.2125
0.3000	1.4237
0.3500	1.6416
0.4000	1.8692
0.4500	2.1095
0.5000	2.3660
0.5500	2.6430
0.6000	2.9462
0.6500	3.2831
0.7000	3.6649
0.7500	4.1083
0.8000	4.6416
0.8500	5.3170
0.9000	6.2514
0.9500	7.8147
0.9800	9.8374
0.9900	11.3449
0.9950	12.8382
0.9980	14.7955
0.9990	16.2662
0.9999	21.1075