UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST 102 – Videregående kurs i statistikk
Eksamensdag:	Mandag 31. mai 1999
Tid til eksamen:	0900 - 1500
Tillatte hjelpemidler:	Formelsamling for ST 101, og ST 102 Karl Rottman: "Mathematische Formelsammlung" Jahren og Knutsen: "Formelsamling i matematikk" Lommeregner.
VEDLEGG :	2 SIDER

Oppgavesettet er på 7 sider.

Kontrollér at oppgavesettet er komplett
før du begynner å besvare spørsmålene.

Oppgave 1

Anta at $X \sim N(\mu, \sigma^2)$, dvs. X har følgende tetthet:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), -\infty < x < \infty$$

(a) Vis at da holder:

$$Z = \frac{X-\mu}{\sqrt{\sigma^2}} \sim N(0, 1)$$

La så S^2 være en estimator for σ^2. Anta videre at S^2 er uavhengig av X og at:

$$Y = \frac{\nu S^2}{\sigma^2} \sim \chi^2_\nu$$

Dvs. tettheten til Y er gitt ved:

$$f_Y(y) = \frac{\frac{1}{2}^{\nu/2}}{\Gamma(\frac{\nu}{2})} y^{(\nu/2)-1} \exp(-y/2), \quad 0 < y < \infty$$

(b) Finn forventning og varians til S^2 og undersøk om estimatoren er konsistent.

(c) Utled tettheten til:

$$T = \frac{X-\mu}{\sqrt{S^2}}$$

Oppgaven fortsetter på neste side

(d) Utled også tettheten til $F = T^2$.

Oppgave 2

Vikarbyrået "Man-Co" ønsker å undersøke om kundene deres er fornøyd med vikarene deres. De foretar derfor en spørreundersøkelse blant $n = 400$ av disse kundene, og ber dem om å angi om de var "godt fornøyd", "sånn passe fornøyd" eller "misfornøyd" med vikarene. Kundene grupperes videre inn bransjevis i kategoriene "handel og kontor", "renhold" og "andre". Tabellen under viser hvordan svarene fordelte seg:

	Handel og kontor	Renhold	Andre
Godt fornøyd	40	59	67
Sånn passe fornøyd	51	53	48
Misfornøyd	35	24	23

La nå for $i, j = 1,2,3$:

$X_{i,j}$ = Antall svar i i-te rad og j-te kolonne.

$p_{i,j}$ = Sannsynligheten for at svaret til en tilfeldig valgt kunde faller i i-te rad og j-te kolonne.

(a) Hva slags fordeling er det rimelig å anta at vektoren $(X_{1,1}, X_{1,2}, X_{1,3}, \ldots, X_{3,3})$ har? Gjør rede for hvilke antagelser som må være oppfylt for at dette skal gjelde. Synes du antagelsene er rimelige? Begrunn svaret.

La nå:

$$p_{i,\cdot} = p_{i,1} + p_{i,2} + p_{i,3}\ , i = 1, 2, 3, \text{ og } p_{\cdot j} = p_{1,j} + p_{2,j} + p_{3,j}\ , j = 1, 2, 3$$

Vi ønsker å teste hypotesen:

$$H_0\colon p_{i,j} = p_{i,\cdot} \cdot p_{\cdot j}\ ,\ i, j = 1,2,3$$

mot alternativet:

$$H_1\colon p_{i,j} \neq p_{i,\cdot} \cdot p_{\cdot j}\ ,\ \text{for minst én } i \text{ og } j.$$

(b) Gi en praktisk fortolkning av H_0 og H_1.

(c) Ved testing av nevnte hypotese, skal det benyttes et signifikansnivå på 0.05. Forklar kort hva dette betyr.

For å teste hypotesen, benyttes følgende testobservator:

Oppgaven fortsetter på neste side

$$C = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\left(X_{i,j} - n\hat{p}_{i,\cdot}\hat{p}_{\cdot,j}\right)^2}{n\hat{p}_{i,\cdot}\hat{p}_{\cdot,j}}$$

der vi har innført estimatorene:

$$\hat{p}_{i,\cdot} = \frac{1}{n}\sum_{j=1}^{3} X_{i,j}\ , \text{ og } \hat{p}_{\cdot,j} = \frac{1}{n}\sum_{i=1}^{3} X_{i,j}$$

(d) Hvilken fordeling har C (tilnærmet) under H_0? For hvilke verdier av C er det rimelig å forkaste nullhypotesen? Begrunn svaret.

For å teste hypotesen benytter vi MINITAB-kommandoen:

Stat > Tables > Chisquare Test...

Resultatet er gitt nedenfor:

```
Chi-Square Test

Expected counts are printed below observed counts

      HandKont  Renhold    Andre    Total
  1        40       59       67      166
        52.29    56.44    57.27

  2        51       53       48      152
        47.88    51.68    52.44

  3        35       24       23       82
        25.83    27.88    28.29

Total     126      136      138      400

ChiSq =  2.889 +  0.116 +  1.653 +
         0.203 +  0.034 +  0.376 +
         3.255 +  0.540 +  0.989 = 10.055
df = 4, p = 0.040
```

(e) Forklar kort hva denne utskriften betyr. Gir dette grunnlag for å forkaste nullhypotesen med det oppgitte signifikansnivået?

"Man-Co" ønsker videre å finne ut om det er slik at deres kunder innen handel og kontor er mindre fornøyd enn kundene i renholdsbransjen.

For å se nærmere på dette, innfører vi:

n_H = Antall kunder innen handel og kontor.

n_R = Antall kunder i renholdsbransjen.

X_H = Antall kunder innen handel og kontor som ikke er misfornøyde.

Oppgaven fortsetter på neste side

X_R = Antall kunder i renholdsbransjen som ikke er misfornøyde.

p_H = Sannsynligheten for at en kunde innen handel og kontor ikke er misfornøyd.

p_R = Sannsynligheten for at en kunde i renholdsbransjen ikke er misfornøyd.

Av tabellen over finner vi at:

$$n_H = 126, \quad n_R = 136, \quad X_H = 91, \quad X_R = 112$$

I denne delen av oppgaven betrakter vi n_H og n_R som gitte (deterministiske) tall. Videre antar vi at X_H og X_R er stokastisk uavhengige, og at:

$$X_H \sim \text{Bin}(n_H, p_H), \quad \text{og} \quad X_R \sim \text{Bin}(n_R, p_R)$$

(f) Foreslå estimatorer for p_H og p_R, og beregn disse med de oppgitte verdier. Hva blir estimatorenes varianser?

Vi stiller så opp følgende hypotese:

$$H_0: p_H = p_R \quad \text{mot} \quad H_1: p_H \neq p_R$$

og innfører testobservatoren:

$$Z = \frac{\frac{X_H}{n_H} - \frac{X_R}{n_R}}{\sqrt{\left(\frac{n_H + n_R}{n_H n_R}\right)\left(\frac{X_H + X_R}{n_H + n_R}\right)\left(1 - \frac{X_H + X_R}{n_H + n_R}\right)}}$$

(g) Gi en begrunnelse for at Z er tilnærmet $N(0, 1)$-fordelt under H_0. For hvilke verdier er det rimelig å forkaste nullhypotesen?

(h) Beregn Z og finn ut om nullhypotesen bør forksates. Benytt 0.05 som signifikansnivå.

(i) Utled et tilnærmet 95%-konfidensintervall for $(p_H - p_R)$, og beregn dette intervallet.

Oppgave 3

Vi skal i denne oppgaven studere hvordan levetiden til en bestemt type teknisk komponent avhenger av temperaturen. For å belyse dette, måler vi levetiden (i timer) til $n = 20$ komponenter. Hver komponent testes i et miljø med konstant temperatur gjennom hele dens levetid. Komponent nr. 1 testes med temperatur 25 °C, komponent nr. 2 testes med temperatur 30 °C, og så videre opp til komponent nr. 20 som testes med temperatur 120 °C.

Vi innfører så for $i = 1, \ldots, 20$:

Oppgaven fortsetter på neste side

X_i = Levetid for i-te komponent.

t_i = Temperatur for i-te komponent

og antar at vi har følgende modell:

$$X_i = a + bt_i + \varepsilon_i, \qquad i = 1,\dots, 20$$

der $\varepsilon_1,\dots, \varepsilon_{20}$ antas å være uavhengige, identisk $N(0, \sigma^2)$-fordelte.

(a) Utled minste kvadraters estimatorer, $\hat{a}$ og $\hat{b}$ for henholdsvis a og b.

[Svar:

$$\hat{a} = \bar{X} - \hat{b}\bar{t}\text{, og } \hat{b} = \frac{1}{M}\sum_{i=1}^{n} X_i(t_i - \bar{t}),\text{ der } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \text{ og } M = \sum_{i=1}^{n}(t_i - \bar{t})^2.]$$

(b) Vis at $\hat{a}$ og $\hat{b}$ blir forventningsrette, og beregn også Var($\hat{a}$), Var($\hat{b}$) og Cov($\hat{a}$, $\hat{b}$).

[Svar:

$$\mathrm{Var}(\hat{a}) = \frac{\sigma^2}{nM}\sum_{i=1}^{n} t_i^2,\ \mathrm{Var}(\hat{b}) = \frac{\sigma^2}{M},\ \mathrm{Cov}(\hat{a},\hat{b}) = -\bar{t}\frac{\sigma^2}{M}.]$$

Vi innfører så:

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(X_i - \hat{a} - \hat{b}t_i\right)^2$$

(c) Vis at S^2 er en forventningsrett estimator for σ^2.

[Hint: Vis først at vi har følgende:

$$(n-2)S^2 = \sum_{i=1}^{n}(X_i - \mathrm{E}(X_i))^2 - n(\bar{X} - \mathrm{E}(\bar{X}))^2 - M(\hat{b} - \mathrm{E}(\hat{b}))^2.]$$

Det kan vises at S^2 og $\hat{b}$ er stokastisk uavhengige.

(d) Benytt dette til å utlede fordelingen for:

$$\frac{\hat{b} - b}{\sqrt{S^2/M}}$$

[Hint: Du kan uten videre benytte resultater fra Oppgave 1.]

(e) Utled et 95%-konfidensintervall for b.

I tabellen nedenfor har vi listet opp levetiden til de 20 komponentene med tilhørende tempe-

Oppgaven fortsetter på neste side

raturer.

Temperatur °C	Levetid (timer)
25	620
30	607
35	543
40	492
45	558
50	543
55	479
60	544
65	563
70	442
75	406
80	427
85	373
90	356
95	373
100	383
105	321
110	339
115	306
120	302

For å analysere datamaterialet benytter vi MINITAB-kommandoen:

Stat > Regression > Regression...

De viktigste resultatene er gitt nedenfor:

Regression Analysis

```
The regression equation is
Levetid = 689 - 3.32 Temp

Predictor        Coef       Stdev    t-ratio        p
Constant       689.32       21.68      31.79    0.000
Temp          -3.3168      0.2779     -11.94    0.000

s = 35.83       R-sq = 88.8%
```

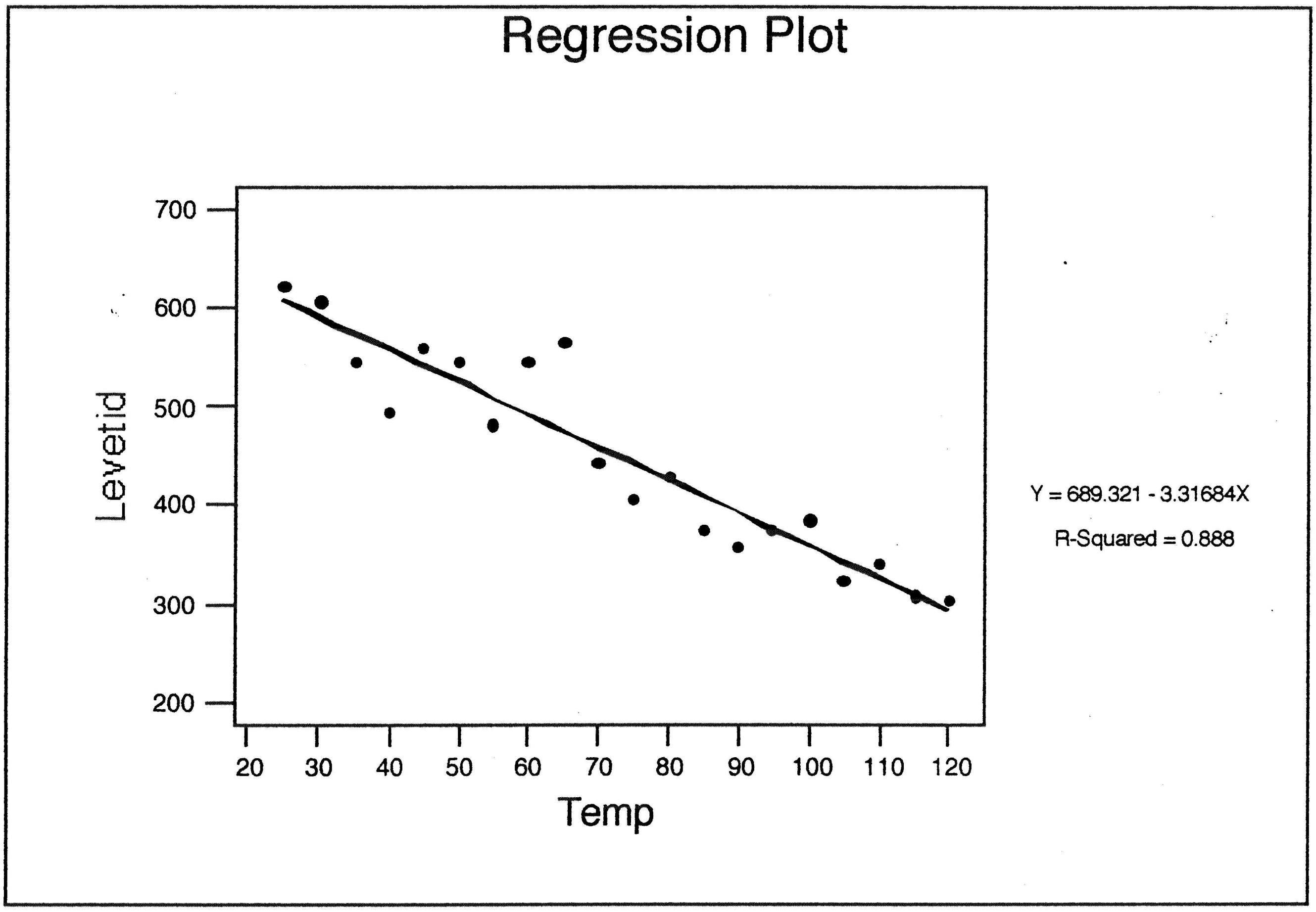
(f) Forklar kort hva denne utskriften betyr. Hvilke konklusjoner kan du trekke ut av dette?

For å få et grafisk bilde av sammenhengen mellom temperatur og levetid benytter vi i tillegg MINITAB-kommandoen:

Oppgaven fortsetter på neste side

Stat > Regression > Fitted Line Plot...

Resultatet blir følgende figur:



(g) Hva kan du lese ut av denne figuren?

SLUTT