TEKNISKA HÖGSKOLAN I LINKÖPING
Matematiska institutionen
Matematisk statistik
EE

Kurskod: TAMS65
Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS
Tentamen tisdagen den 9 juni 2009 kl 14-18.

Hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen samt räknedosa med tömda minnen.

Betygsgränser: 8-11 poäng ger betyg 3, 11.5-14.5 poäng ger betyg 4, 15-18 poäng ger betyg 5.

Jourhavande lärare: Eva Enqvist, tel 281433.

Resultatet meddelas via LADOK.

Tydliga motiveringar krävs till varje uppgift.

1. a) Man har ställt upp hypotesen att det under lördagar och söndagar inträffar dubbelt så många bilolyckor med dödsfall som under övriga veckodagar. Detta skulle innebära att sannolikheten att en slumpmässigt vald olycka av denna kategori inträffar en lördag är 2/9, en söndag 2/9 och övriga veckodagar 1/9. Man har valt 90 olyckor slumpmässigt i en stor databas och fått följande fördelning över veckans dagar:

Sön	Mån	Tis	Ons	Tor	Fre	Lör
29	6	8	11	7	10	19

Motsäger detta resultat den ställda hypotesen? Genomför ett lämpligt χ^2-test på nivån cirka 5%. (1.5p)

b) Utmattning i metaller uppstår genom upprepad påverkan av små laster. För att testa en metall utsätter man provstavar för cykler av belastningar tills brott uppstår. Under ett utmattningstest fick en tillverkare följande data (enhet: cykler $\cdot 10^5$) för tiden till utmattning

4.39 4.75 4.20 4.66 4.72 4.28 4.40 4.48

Modell: Vi har oberoende observationer $x_1, \ldots, x_8$ från $N(\mu, 0.20)$.

Undersök med hjälp av ett lämpligt test eller konfidensintervall om man med någon säkerhet kan hävda att $\mu \neq 4.30$. Nivå 5%. (1.5p)

2. Man har gjort oberoende mätningar av exekveringstider $x_1, \ldots, x_n$ (enhet: s) för $n = 40$ jobb på en datacentral. Resultat:

10	19	90	40	15	11	32	17	4	152	23	13	36	101	2	14	2	23	34	15
27	1	57	17	3	30	50	4	62	48	9	11	20	13	38	54	46	12	5	26

där $\overline{x} = 29.65$.

Modell: De s.v. X_i är exponentialfördelade med väntevärde μ, d.v.s. med täthetsfunktion

$$f(x) = \frac{1}{\mu} e^{-x/\mu} \quad \text{för} \quad x \geq 0.$$

a) Konstruera ett tvåsidigt approximativt 95% konfidensintervall för μ. (2p)

b) Låt g vara en gräns sådan att $P(X_i \le g) = 0.90$. Konstruera med hjälp av resultatet i a) ett approximativt 95% konfidensintervall för g. (1p)

3. Ett amerikanskt företag ville jämföra hållbarheten hos den egna färgen med hållbarheten hos ett konkurrerande märke. Färgens livslängd, y mätt i år bestämdes för olika orter i USA, tre mätningar för varje ort och märke, se nedan.

	Avg. Temp (°F)		Mean Annual	Lifetime (years)	
City	January	July	Precipitation (in.)	Sponsor's Paint	Competitor's Paint
Atlanta, GA	41.9	78.6	48.6	11.5 10.7 12.3	10.8 11.1 10.2
Boston, MA	29.6	73.5	43.8	11.7 10.1 12.5	10.7 11.6 11.0
Kansas City, KS	28.4	80.9	29.3	12.3 13.4 12.8	11.8 12.2 11.3
Minneapolis, MN	11.2	73.1	26.4	10.5 9.9 11.2	10.4 9.6 9.2
Dallas, TX	45.0	86.3	34.2	11.2 10.6 12.0	10.6 10.1 11.4
Denver, CO	29.5	73.3	15.3	15.2 14.2 13.8	13.4 14.4 13.2
Miami, FL	67.1	82.4	57.5	8.7 7.9 9.4	8.1 8.6 7.6
Phoenix, AZ	52.3	92.3	7.1	11.1 11.8 12.4	10.9 10.1 9.9
San Francisco, CA	48.5	62.2	19.7	16.7 17.2 15.9	15.8 15.4 14.9
Seattle, WA	40.6	65.3	38.9	14.2 14.1 13.6	12.6 13.6 14.1
Washington, DC	35.2	78.9	39.0	12.6 11.5 12.0	11.9 10.9 11.4

I tabellen anges också

$x_1 =$ genomsnittstemperatur i ^oF för januari,
$x_2 =$ genomsnittstemperatur i ^oF för juli,
$x_3 =$ genomsnittlig årsnederbörd (enhet: inch).

Vidare låter vi

$$x_4 = \begin{cases} 1 & \text{för konkurrentens färg} \\ 0 & \text{annars} \end{cases}$$

$$x_5 = \begin{cases} 1 & \text{för Miami} \\ 0 & \text{annars} \end{cases}$$

Datorutskrifter från Minitab finns på nästa sida.

a) Vilken är den bästa ensamma förklaringsvariabeln bland $x_1, \ldots, x_4$? Motivera ditt svar kortfattat. (0.5p)

b) Data har analyserats enligt två olika modeller.

Modell 1: $Y = \beta_0' + \beta_1 x_1 + \beta_2' x_2 + \beta_3' x_3 + \beta_4' x_4 + \varepsilon'$
Modell 2: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$
där ε-variablerna antas vara normalfördelade.

Beskriver modell 2 data bättre än modell 1? Motivera ditt svar med hjälp av ett lämpligt test på nivån 0.05. Både nollhypotes och mothypotes ska anges. Slutsatsen av testet ska tydligt framgå. (1.5p)

c) Kan man utgående från analys nr 2 påvisa en systematisk skillnad mellan de båda färgtyperna och därför rekommendera någon av dem? Motivera ditt svar med hjälp av ett lämpligt 95% konfidensintervall. (1p)

Minitabanalyser

```
MTB > corr c1-c5
```

Correlations: x1, x2, x3, x4, Y

```
        x1      x2      x3      x4
x2   0.297

x3   0.246  -0.038

x4   0.000   0.000   0.000

Y   -0.140  -0.632  -0.477  -0.190

Cell Contents: Pearson correlation
```

ANALYS NR 1

```
MTB > Regress 'Y' 4 'x1'-'x4';
SUBC>   Constant;
SUBC>   Brief 2.
```

Regression Analysis: Y versus x1, x2, x3, x4

```
The regression equation is
Y = 27.5 + 0.0310 x1 - 0.178 x2 - 0.0811 x3 - 0.794 x4


Predictor       Coef   SE Coef
Constant      27.450     1.361
x1           0.03098   0.01083
x2          -0.17842   0.01770
x3          -0.08106   0.01019
x4           -0.7939    0.2804


S = 1.13893   R-Sq = 72.4%


Analysis of Variance

Source          DF        SS      MS
Regression       4   207.814  51.953
Residual Error  61    79.127   1.297
Total           65   286.941
```

ANALYS NR 2

```
MTB > Name c8 "FITS1"
MTB > Regress 'Y' 5 'x1'-'x4' 'x5' ;
SUBC>   Fits 'FITS1';
SUBC>   Constant;
SUBC>   Brief 2.
```

Regression Analysis: Y versus x1, x2, x3, x4, x5

```
The regression equation is
Y = 24.7 + 0.0671 x1 - 0.170 x2 - 0.0509 x3 - 0.794 x4 - 3.51 x5


Predictor        Coef    SE Coef
Constant       24.707      1.184
x1            0.06713    0.01060
x2           -0.16976    0.01424
x3          -0.050917   0.009613
x4            -0.7939     0.2245
x5            -3.5142     0.5929


S = 0.911998   R-Sq = 82.6%


Analysis of Variance

Source          DF        SS      MS
Regression       5   237.037  47.407
Residual Error  60    49.904   0.832
Total           65   286.941
```

4. Om en stormarknad ofta har extrapris på en vara så anses det att kunderna förväntar sig ett lägre pris.

I en amerikansk managementkurs fick studenter information om priset på en vara på tre olika stormarknader under tio veckor. De hade slumpmässigt delats upp i grupper med elva studenter i varje. Varje student följde prisutvecklingen för en stormarknad och fick sedan uppge ett pris som de väntade sig för den elfte veckan. Normalpriset var detsamma för alla tre stormarknaderna. Studenterna angav sina priser oberoende av varandra.

För varje stormarknad hade man också noterat antalet extrapriskampanjer under den aktuella tidsperioden.

Resultat (studenternas priser för den elfte veckan):

```
TRE EXTRAPRISKAMPANJER:
 C1
   4.12  3.91  3.96  4.22  3.88  4.14  4.17  4.07  4.16  4.12  3.84
FEM EXTRAPRISKAMPANJER:
 C2
   3.32  3.86  3.65  3.71  3.78  3.93  3.73  3.71  3.69  3.83  3.99
SJU EXTRAPRISKAMPANJER:
 C3
   3.45  3.64  3.37  3.27  3.58  3.67  3.74  3.50  3.60  3.57  3.50
```

SAMMANFATTNING AV DATA

	N	MEAN	STDEV
C1	11	4.0536	0.1320
C2	11	3.7555	0.1939
C3	11	3.5355	0.1363

Modell: Vi har tre oberoende stickprov från $N(\mu_i, \sigma)$.

a) Konstruera ett uppåt begränsat 95% konfidensintervall för σ (1p)

b) Har man olika förväntningar på priset för de tre stormarknaderna? Motivera ditt svar genom att konstruera lämpliga konfidensintervall vart och ett med konfidensgraden 99%.

Verkar antalet extrapriskampanjer påverka priset så som man misstänkte? Motivera ditt svar kortfattat med hjälp av dina konfidensintervall. (2p)

5. Ett irritationsmoment vid kopiering är att papper kan fastna i maskinen. Efter service av maskinen garanterar firman att frammatningen fungerar med sannolikhet 0.90 i varje enskild kopiering av ett papper.

Man vill pröva

$$H_0 : p = 0.9 \quad \text{mot} \quad H_1 : p < 0.9$$

på nivån 0.05. Man ska genomföra provkopieringar och räkna hur många gånger papperet fastnat/inte fastnat. Hur många oberoende kopieringar av ett papper behöver man åtminstone göra, om man vill att testets styrka för $p = 0.8$ ska vara minst 0.95. Lämplig approximation får utnyttjas. (3p)

6. Livslängder i år för en viss sorts verktyg kan anses vara stokastiska variabler med fördelningsfunktion

$$F(x) = 1 - (1 + x)^\lambda e^{-\lambda x} \ ; \text{ för } \quad x > 0\,,$$

där $\lambda > 0$. Alltså är täthetsfunktionen

$$f(x) = \lambda(1 + x)^\lambda e^{-\lambda x}(1 - \frac{1}{1+x}) \quad \text{för} \quad x > 0.$$

a) Man har observerat oberoende livslängder $x_1, \ldots, x_n$ för sådana verktyg. Härled ML-skattningen av λ. (1.5p)

b) Av 100 sådana här verktyg fungerade 85 fortfarande efter ett års användning. Punktskatta parametern λ med hjälp av denna information. (1.5p)

Lösningar till tentamen i TAMS65, 2009-06-09.

1a) $H_0: P(\text{lör}) = P(\text{sön}) = \frac{2}{9}$, $P(\text{mån}) = \ldots = P(\text{fre}) = \frac{1}{9}$.

H_1: H_0 ej sann

Vi får de förväntade frekvenserna, $90p_i$:

Sön	Mån	Tis	Ons	Tor	Fre	Lö
20	10	10	10	10	10	20

Teststorhet: $Q = \sum_1^7 \frac{(N_i - 90p_i)^2}{90p_i} = 7.10$

H_0 förkastas om $Q > c$. Den s.v. Q är appr $\chi^2(7-1)$ om H_0 är sann. Tabell ger $c = 12.60$.

$7.10 < 12.60$. H_0 kan inte förkastas. Utgående från givna data finns det ingen anledning att förkasta H_0.

b) $H_0: \mu = 4.30$ mot $H_1: \mu \neq 4.30$; nivå 5%.

Teststorhet: $u = \frac{\bar{x} - 4.30}{0.20/\sqrt{8}} = \frac{4.485 - 4.30}{0.20/\sqrt{8}} = 2.62$

Den s.v. $U \sim N(0,1)$ om H_0 är sann. H_0 förkastas om $u < c$ eller $u > c$. $\Phi(c) = 0.975$ ger $c = 1.96$.

$2.62 > 1.96$; H_0 förkastas. Med stor sannolikhet är $\mu \neq 4.30$.

2a) $\hat{\mu} = \bar{x} = 29.65$

Centrala gränsvärdessatsen ger att den s.v.

$\bar{X}$ appr $N(\mu, \frac{\mu}{\sqrt{40}})$

Hjälpvariabel $\frac{\bar{X} - \mu}{\mu/\sqrt{40}}$ appr $N(0,1)$.

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\mu/\sqrt{40}} < 1.96\right) \simeq 0.95$$

$$P\left(-\frac{1.96}{\sqrt{40}} < \frac{\bar{X}}{\mu} - 1 < \frac{1.96}{\sqrt{40}}\right) \simeq 0.95$$

$$P\left(1-\frac{1.96}{\sqrt{40}} < \frac{\bar{X}}{\mu} < 1+\frac{1.96}{\sqrt{40}}\right) \simeq 0.95$$

$$P\left(\frac{\bar{X}}{1+1.96/\sqrt{40}} < \mu < \frac{\bar{X}}{1-1.96/\sqrt{40}}\right) \simeq 0.95$$

Ger $I_\mu = (22.64, 42.96)$

(Även $I_\mu = (\hat{\mu} \mp 1.96 \frac{\hat{\mu}}{\sqrt{40}})$ är OK.)

b) $P(X \leq g) = 0.90 \Leftrightarrow \int_0^g \frac{1}{\mu} e^{-x/\mu} dx = 0.90$

$1 - e^{-g/\mu} = 0.90 \Leftrightarrow e^{-g/\mu} = 0.10 \Leftrightarrow$

$g = -\mu \ln 0.10 = 2.3026 \mu$

Detta ger $I_g = (2.3026 \cdot 22.64, 2.3026 \cdot 42.96) \simeq$

$\simeq (52.1, 98.9)$

3a) Den bästa ensamma förkl. variabeln bland $x_1, \ldots, x_4$ är den som är starkast korrelerad med y, d.v.s. x_2.

b) $H_0: \beta_5 = 0$ (modell 1 duger)

$H_1: \beta_5 \neq 0$ (modell 2 bättre)

prövas med hjälp av $W = \dfrac{(Q_{res}^{(1)} - Q_{res}^{(2)})/1}{Q_{res}^{(2)}/60} = 35.14$

H_0 förkastas om $W > c$. Den s.v. $W \sim F(1,60)$ om H_0 är sann. Tabell ger $c \approx 4.01$.

$35.14 > 4.01$; H_0 förkastas. Modell 2 beskriver data bättre, vilket tyder på att mätningarna för Miami avviker från de andra.

c) $$E(Y) = \begin{cases} \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_5 x_5 & \text{egna färg.} \\ \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 + \beta_5 x_5 & \text{kork. -"-} \end{cases}$$

Systematisk skillnad: β_4

$I_{\beta_4} = (\hat{\beta}_4 \mp t \cdot s\sqrt{h_{44}}) = (-0.7939 \mp 2.00 \cdot 0.2245) =$

$= (-1.24, -0.34)$ där $t = 2.00$ ges i $t(60)$-tabell.

Bara negativa värden. Konkurrentens färg verkar ha kortare livslängd och man kan rekommendera den egna färgen.

4a) $s^2 = \dfrac{10s_1^2 + 10s_2^2 + 10s_3^2}{30} = 0.02453$

$s = 0.1566$; frihetsgrad: 30

Den s.v. $30S^2/\sigma^2 \sim \chi^2(30)$

$P(18.49 < \dfrac{30S^2}{\sigma^2}) = 0.95$

$P(\sigma^2 < 30S^2/18.49) = 0.95$

$I_\sigma = (0, s\sqrt{30/18.49}) = (0, 0.200)$

b) Vi konstruerar intervallen $I_{\mu_i - \mu_j}$.

$\hat{\mu}_i - \hat{\mu}_j = \bar{x}_i - \bar{x}_j$; den s.v. $\bar{X}_i - \bar{X}_j \sim N(\mu_i - \mu_j, \sigma\sqrt{\frac{2}{11}})$

Hjälpvariabeln $\dfrac{\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)}{S\sqrt{2/11}} \sim t(30)$ och ger

$I_{\mu_i - \mu_j} = (\bar{x}_i - \bar{x}_j \mp t \cdot s \cdot \sqrt{2/11}) = (\bar{x}_i - \bar{x}_j \mp 0.1836)$

där $t = 2.75$

$I_{\mu_1 - \mu_2} = (0.114, 0.482)$; $I_{\mu_1 - \mu_3} = (0.334, 0.702)$

$I_{\mu_2 - \mu_3} = (0.036, 0.404)$.

Vi ser att $\mu_1 > \mu_2 > \mu_3$ med stor sannolikhet, d.v.s. ju fler extrapriskampanjer desto lägre förväntat pris.

5) Man gör n provkopieringar.

x = antalet ggr papperet inte fastnar

Teststorhet: x

Den s.v. $X \sim Bin(n, p)$ appr $N(np, \sqrt{np(1-p)})$

om $np(1-p) > 10$.

H_0 förkastas om $x \le a$.

$$\begin{cases} 0.05 = P(X \le a \text{ om } p = 0.9) \approx \Phi\left(\frac{a-0.9n}{\sqrt{0.09n}}\right) \\ 0.95 = P(X \le a \text{ om } p = 0.8) \approx \Phi\left(\frac{a-0.8n}{\sqrt{0.16n}}\right) \end{cases}$$

$$\begin{cases} \frac{a-0.9n}{\sqrt{0.09n}} = -1.645 \\ \frac{a-0.8n}{\sqrt{0.16n}} = 1.645 \end{cases} \text{ ger } \begin{cases} a-0.9n = -1.645 \cdot 0.3\sqrt{n} & (1) \\ a-0.8n = 1.645 \cdot 0.4\sqrt{n} & (2) \end{cases}$$

(2)-(1) ger $0.1n = 1.645 \cdot 0.7\sqrt{n}$ d.v.s.

$\sqrt{n} = 16.45 : 0.7$; $n = 132$

Man bör kopiera minst 133 ggr.

Kontroll: $133 \cdot 0.1 \cdot 0.9 = 11.97 > 10$ och $133 \cdot 0.8 \cdot 0.2 = 21.28 > 10$. Normalappr. tillåten.

6a) $L(\lambda) = \prod_{i=1}^{n} \left[\lambda(1+x_i)^\lambda e^{-\lambda x_i}\left(1-\frac{1}{1+x_i}\right)\right] =$

$= \lambda^n \left[\prod_{i=1}^{n}(1+x_i)^\lambda\right] e^{-\lambda \sum_1^n x_i} \left[\prod_{i=1}^{n}\left(1-\frac{1}{1+x_i}\right)\right]$

$\ell(\lambda) = \ln L(\lambda) = n\ln\lambda + \sum_{i=1}^{n}\lambda\ln(1+x_i) - \lambda\sum_1^n x_i + \text{konst}$

$\ell'(\lambda) = \frac{n}{\lambda} + \sum_1^n \ln(1+x_i) - \sum_1^n x_i$

$\ell'(\lambda) = 0$ för $\hat{\lambda} = \frac{n}{\sum_1^n (x_i - \ln(1+x_i))}$

$\ell''(\lambda) = -\frac{n}{\lambda^2} < 0$; alltså är $\hat{\lambda}$ ML-skattn.

Eftersom $\ln(1+x_i) < x_i$ då $x_i > 0$, är $\hat{\lambda} > 0$.

b) $y = 85$ är obs. av $Y \sim Bin(100, p)$, där

$p = P(X_i > 1) = 1 - P(X_i \le 1) = 1 - F(1) = 2^\lambda e^{-\lambda} = \left(\frac{2}{e}\right)^\lambda$

$\hat{p} = 0.85$ ger $\left(\frac{2}{e}\right)^{\hat{\lambda}} = 0.85$ d.v.s.

$\hat{\lambda} = \frac{\ln 0.85}{\ln 2 - \ln e} \approx 0.53$