TEKNISKA HÖGSKOLAN I LINKÖPING
Matematiska institutionen
Matematisk statistik
EE

TAMS 65 MATEMATISK STATISTIK I FORTSÄTTN.KURS
Tentamen fredagen den 7 mars 2008 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.

Obs! Skriv namn och personnummer på varje inlämnat papper.

Observera att uppgift 1 ger 2p och uppgift 3 ger 4p.

1. Ett marknadsinstitut genomför intervjuundersökningar för att bedöma olika produkters marknadspotential. Man väljer slumpmässigt ut personer ur intressanta åldersgrupper, kontaktar dem och ber att få intervjua dem. Bland 200 tillfågade män och 300 tillfrågade kvinnor valde 120 respektive 210 att låta sig intervjuas. Låt p_1 och p_2 vara sannolikheterna att en man respektive en kvinna deltar i en intervju av den aktuella typen. Kan vi med någon säkerhet påstå att det finns skillnad mellan p_1 och p_2? Konstruera ett lämpligt konfidensintervall med konfidensgraden approximativt 95%. Redovisa Din slutsats av intervallet i ord. Personerna fattar beslut om intervju oberoende av varandra. (2p)

2. På en fabrik finns ett ganska stort antal maskiner och man har en reparationsavdelning som sköter reparationerna för samtliga maskiner. Reparatör anropas så snart behov uppstår. Man har under en tidsperiod noterat hundra oberoende väntetider $x_1, \ldots, x_{100}$ (enhet:timme) mellan anrop av reparatör under arbetstid. Vidare har man beräknat $\bar{x} = 0.671$, gjort en klassindelning av mätvärdena och räknat antalet observationer i intervallen:

$0 < x_i \leq 0.4$	$0.4 < x_i \leq 0.8$	$0.8 < x_i \leq 1.2$	$1.2 < x_i$
42	32	13	13

(För att underlätta beräkningarna har vi här använt onödigt få klasser.)

a) Undersök med hjälp av ett lämpligt χ^2- test om det verkar rimligt att anta att de s.v. X_i är exponentialfördelade med ett okänt väntevärde μ. Nivå ungefär 0.05. (2.5p)

b) Reparation av en maskin tar i genomsnitt 1.25 timmar. Hur många reparatörer bedömer du att man bör ha? Motivera ditt svar kortfattat. (0.5p)

3. I datamaterialet nedan finns uppgifter för två olika tidskrifter om antalet helsidesannonser, x, och det totala antalet sidor, Y, för fyra olika utgåvor under 1989 ($t = 1$), 1991 ($t = 2$) och 1993 ($t = 3$). För att kunna skilja på de båda tidskrifterna har tidskrift A kodats med $u = 0$ och tidskrift B med $u = 1$.

	C1	C2	C3	C4
	t	x	u	Y
1	1	60	0	76
2	1	61	0	76
3	1	45	0	84
4	1	72	0	100
5	2	36	0	74
6	2	63	0	74
7	2	34	0	68
8	2	60	0	90
9	3	27	0	76
10	3	53	0	82
11	3	36	0	76
12	3	61	0	90
13	1	53	1	88
14	1	56	1	122
15	1	33	1	88
16	1	44	1	106
17	2	32	1	82
18	2	58	1	104
19	2	55	1	96
20	2	64	1	140
21	3	75	1	202
22	3	58	1	90
23	3	45	1	86
24	3	35	1	86

Vid en preliminär analys visade det sig att t inte gjorde någon större nytta som förklaringsvariabel, så du ska använda modellen

$$Y = \beta_0 + \beta_1 x + \beta_2 u + \varepsilon$$

där $\varepsilon \sim N(0, \sigma)$.

a) Visa hur $R^2 =$R-sq, som finns i datorutskriften på nästa sida, beräknas och förklara kortfattat vad den mäter. (0.5p)

b) Verkar det finnas en systematisk skillnad i antalet sidor mellan de båda tidskrifterna, en skillnad som inte förklaras av antalet helsidesannonser? Besvara frågan med hjälp av ett lämpligt 95% konfidensintervall och redovisa din slutsats beträffande den eventuella skillnaden. (1.5p)

c) Låt oss anta att modellen ovan kan tillämpas för 1994 och att man för tidskrift B till det första numret har fått beställning på 60 helsidesannonser. Konstruera ett lämpligt 95% intervall som beskriver antalet sidor i detta nummer av tidskriften. Är du nöjd med intervallet? (2p)

Datorutskrift till uppgift 3.

```
MTB > Name m1 = 'XPXI1'
MTB > Regress 'Y' 2 'x' 'u';
SUBC>   XPXInverse 'XPXI1';
SUBC>   Constant;
SUBC>   Brief 2.
Regression Analysis: Y versus x, u

The regression equation is

Y = 23.5 + 1.13 x + 27.0 u
```

Predictor	Coef	SE Coef (Stdev)
Constant	23.45	16.78
x	1.1259	0.3107
u	27.000	8.233

S = 20.17 R-Sq = 53.2%

Analysis of Variance

Source	DF	SS	MS
Regression	2	9715.5	4857.8
Residual Error	21	8540.5	406.7
Total	23	18256.0	

Data Display

Matrix XPXI1

$$\begin{matrix} 0.692616 & -0.012025 & -0.083333 \\ -0.012025 & 0.000237 & 0.000000 \\ -0.083333 & 0.000000 & 0.166667 \end{matrix} = (\mathbf{X}^T\mathbf{X})^{-1}$$

4. En tvättmedelstillverkare vill i en reklamkampanj kunna säga att ett av de fyra tvättmedel företaget tillverkar i genomsnitt ger vitare tvätt än de övriga. Man valde 44 personer och delade slumpmässigt in dem i fyra grupper med 11 personer för varje tvättmedel. Varje person genomförde fyra tvättar med ett tvättmedel och poägsatte vitheten. För varje person beräknades sedan en poängsumma för de fyra tvättomgångarna. Höga värden innebär vitare tvätt. Resultat:

Row	x1	x2	x3	x4
1	32.0	31.4	30.9	33.1
2	34.5	32.6	29.4	36.8
3	29.5	35.1	33.1	36.2
4	32.2	36.8	32.7	34.5
5	28.7	29.5	31.5	35.9
6	28.9	33.4	29.6	35.4
7	30.9	35.2	35.3	34.2
8	31.4	30.7	31.2	31.4
9	27.1	34.9	32.0	29.8
10	30.2	33.1	28.5	34.7
11	29.9	30.2	31.4	33.6

Även stickprovsmedelvärde $\bar{x}_i$ och stickprovsstandardavvikelse s_i beräknades för de fyra stickproven:

Level	N	Mean	StDev
x1	11	30.482	2.023
x2	11	32.991	2.357
x3	11	31.418	1.900
x4	11	34.145	2.102

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

a) Kan man med någon säkerhet hävda att $\sigma < 2.6$? Motivera ditt svar med hjälp av ett lämpligt 95% konfidensintervall. (1p)

b) Kan man med någon säkerhet hävda att något av tvättmedlen i ger vitare tvätt än de övriga? Motivera ditt svar med hjälp av lämpliga tvåsidiga konfidensintervall vart och ett med konfidensgrad 98%. Du behöver inte skriva ut alla intervallen, men det ska tydligt framgå hur du drar dina slutsatser. (2p)

5. En liten möbelfabrik tillverkar bord av en speciell typ. Man har noterat en uppgång i antalet beställda bord och måste eventuellt anställa ytterligare en person. Under den senaste månaden har man fått in 32 beställningar av det aktuella bordet.

Modell: Antalet beställningar under t månader är en s.v. $X \sim Po(\lambda t)$.

Om $\lambda > 25$ anser man sig behöva mer personal.

a)Pröva på nivån 0.05 hypotesen

$$H_0 : \lambda = 25 \quad \text{mot} \quad H_1 : \lambda > 25 . \qquad (1.5p)$$

b)Under hur många månader framåt i tiden behöver man räkna beställningar om man vill genomföra hypotesprövningen i a) och man vill att styrkan för testet ska vara minst 0.90 om $\lambda = 35$. Observationen för den senaste månaden ska alltså inte användas i det nya testet. (1.5p)

Lämpliga approximationer får utnyttjas både i a) och b).

6. En liten öppenvårdsmottagning sköts av två läkare. De har var och en betjäningsintensiteten 3 patienter per timme. Under den tid mottagningen är öppen anländer patienter enligt en Poissonprocess med intensiteten λ patienter per timme. Alla patienter antas få plats i väntrummet. För att få rimliga köer måste man ha en viss överkapacitet och man har bedömt att man helst ska $\lambda < 5$. På senare tid har köerna ökat, så man misstänker att patientströmmen har ökat. Vid tio oberoende tillfällen har man räknat antalet patienter på mottagningen inklusive de patienter som håller på att bli behandlade och fått observationer, x_i, där $i = 1, \ldots, 10$:

2 14 5 10 12 7 4 4 6 9

Man kan visa (men det behöver du inte göra) att vid stationaritet gäller att sannolikhetsfunktionen

$$P(X_i = x) = \begin{cases} \dfrac{6-\lambda}{6+\lambda} & \text{om } x = 0 \\ \dfrac{6-\lambda}{6+\lambda} \cdot 2\left(\dfrac{\lambda}{6}\right)^x & \text{om } x = 1, 2, \ldots \end{cases}$$

dät $\lambda < 6$.

a) Bestäm ML-skattningen av λ. (2p)

b) Uppskatta med hjälp av resultatet i a) sannolikheten att det finns minst åtta patienter på mottagningen, inklusive de patienter som bahandlas. (1p)

Du får i både a) och b) anta att stationaritet råder.

Anm. Det finns effektivare sätt att uppskatta λ, om man gör en annan typ av mätningar.

Lösningar till tentamen i TAMS 65, 2008-03-07.

1. $x = 120$ är observation av $X \sim Bin(n_1, p_1)$; $n_1 = 200$.

$y = 210$ är observation av $Y \sim Bin(n_2, p_2)$; $n_2 = 300$.

$\hat{p}_1 = \frac{x}{n_1} = 0.6$; $\hat{p}_2 = \frac{y}{n_2} = 0.7$;

$\hat{p}_2 - \hat{p}_1 = \frac{y}{n_2} - \frac{x}{n_1} = 0.1$

De s.v. X och Y är appr. normalfördelade, eftersom $n_1\hat{p}_1(1-\hat{p}_1) > 10$ och $n_2\hat{p}_2(1-\hat{p}_2) > 10$. Då är även den s.v. $\hat{P}_2 - \hat{P}_1$ appr. normalförd. med parametrar

$$E(\hat{P}_2 - \hat{P}_1) = \frac{1}{n_2}E(Y) - \frac{1}{n_1}E(X) = p_2 - p_1$$

$$Var(\hat{P}_2 - \hat{P}_1) = Var\left(\frac{Y}{n_2}\right) + Var\left(\frac{X}{n_1}\right) =$$

$$= \frac{1}{n_2^2}\cdot n_2 p_2(1-p_2) + \frac{1}{n_1^2}\cdot n_1 p_1(1-p_1) = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$$

Hjälpvariabeln $\dfrac{\hat{P}_2 - \hat{P}_1 - (p_2 - p_1)}{\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}}$ appr $N(0,1)$

och den ger

$$I_{p_2-p_1} = \left(\hat{p}_2 - \hat{p}_1 \mp 1.96\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}\right)$$

$$= (0.1 \mp 0.085) = (0.015,\ 0.185)$$

Bara positiva värden i $I_{p_2-p_1}$. Alltså är $p_2 > p_1$ med stor slh, dvs kvinnor är mera benägna att delta i intervjuer av den aktuella typen.

2a) H_0: Vi har observationer från exponentialfördelning med väntevärde μ.

$\hat{\mu} = \bar{x} = 0.671$

Vi beräknar först sannolikheterna för

intervallen då H_0 är sann och approximativa värden på dem.

$$p_1 = \int_0^{0.4} \frac{1}{\mu} e^{-x/\mu} dx = 1 - e^{-0.4/\mu}$$

P s s får vi

$$p_2 = e^{-0.4/\mu} - e^{-0.8/\mu}$$

$$p_3 = e^{-0.8/\mu} - e^{-1.2/\mu}$$

$$p_4 = e^{-1.2/\mu}$$

$$p_1 \simeq 1 - e^{-0.4/\hat{\mu}} = 1 - 0.5509 = 0.4491$$

$$p_2 \simeq 0.5509 - 0.3035 = 0.2474$$

$$p_3 \simeq 0.3035 - 0.1672 = 0.1363$$

$$p_4 \simeq 0.1672$$

De förväntade frekvenserna $100p_j$ är ungefär

44.91 24.74 13.63 16.72

$$Q = \sum_1^4 \frac{(N_j - 100p_j)^2}{100p_j} \simeq 3.176$$

H_0 förkastas om $Q > c$. Den s.v. Q är appr $\chi^2(4-1-1)$ om H_0 är sann. Tabell ger $c = 5.99$.

$Q = 3.176 < 5.99$; H_0 kan inte förkastas.

De s.v. X_i kan vara exponentialfördelade.

b) Man har i genomsnitt $\frac{1}{0.671} = 1.49$ anrop av reparatör per timme.

En reparatör klarar i genomsnitt $\frac{1}{1.25} = 0.8$ reparationer per timme.

Man behöver minst två reparatörer.

3a) $R^2 = Q_{REGR} / Q_{TOT} = 9715.5/18256.0 = 53.22\%$

visar hur stor andel av variationerna hos Y, som förkaras av x och u.

b) $E(Y) = \beta_0 + \beta_1 x$ för tidskrift A

$E(Y) = \beta_0 + \beta_1 x + \beta_2$ för tidskrift B

Parametern β_2 beskriver den systematiska skillnaden i antal sidor mellan B och A.

$$I_{\beta_2} = (\hat{\beta}_2 \mp t \cdot s\sqrt{h_{22}}) = (27.000 \mp 2.08 \cdot 8.233) =$$

$$= (27.000 \mp 17.125) \approx (9.9,\ 44.1)$$

där $t = 2.08$ ges i $t(21)$-tabell.

Bara positiva värden i I_{β_2}. Tidskrift B har i genomsnitt flera sidor än A.

c) Vi söker ett prediktionsintervall för

$$Y_0 = \beta_0 + 60\beta_1 + \beta_2 + \varepsilon_0 = \mu_0 + \varepsilon_0$$

$$\mu_0 = (1 \ \ 60 \ \ 1)\beta$$

$$\hat{\mu}_0 = (1 \ \ 60 \ \ 1)\hat{\beta} = 118.004$$

Den s.v. $Y_0 - \hat{\mu}_0$ är normalfördelad med väntevärde 0 och varians

$$Var(Y_0 - \hat{\mu}_0) = Var(Y_0) + Var(\hat{\mu}_0) =$$

$$= \sigma^2 + (1 \ 60 \ 1)\,\sigma^2 \cdot (X^T X)^{-1} \begin{pmatrix} 1 \\ 60 \\ 1 \end{pmatrix} = \sigma^2 \cdot 1.103$$

σ^2 skattas med $s^2 = Q_{RES}/21$; $s = 20.17$; fr.gr.: 21

Hjälpvariabeln $\dfrac{Y_0 - \hat{\mu}_0}{s\sqrt{1.103}} \sim t(21)$ och den ger

$$I_{Y_0} = (\hat{\mu}_0 \mp t \cdot s\sqrt{1.103}) = (118.004 \mp 44.061) \simeq$$

$$\simeq (74,\ 162)$$

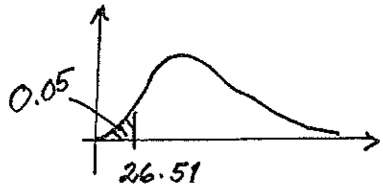
Långt! Modellen beskriver inte data så bra, vilket vi också såg på förklaringsgraden.

4a) Vi konstruerar ett uppåt begränsat konfidensintervall för σ.

σ^2 skattas med $s^2 = \dfrac{10s_1^2 + 10s_2^2 + 10s_3^2 + 10s_4^2}{40} =$

$= 4.419$; $s = 2.102$; fr.gr.: 40

Den s.v. $\dfrac{40S^2}{\sigma^2} \sim \chi^2(40)$



$P\left(26.51 \le \dfrac{40S^2}{\sigma^2}\right) = 0.95$

$P\left(\sigma^2 \le \dfrac{40S^2}{26.51}\right) = 0.95$

$I_\sigma = \left(0, s\sqrt{\dfrac{40}{26.51}}\right) = (0, 2.58)$

Ja, $\sigma \le 2.58 < 2.60$ med stor sannolikhet.

b) Vi konstruerar intervallen $I_{\mu_i - \mu_j}$ vart och ett med konfidensgrad 0.98.

$\hat{\mu}_i - \hat{\mu}_j = \bar{x}_i - \bar{x}_j$

Den s.v. $\bar{X}_i - \bar{X}_j \sim N\left(\mu_i - \mu_j, \sqrt{\dfrac{\sigma^2}{11} + \dfrac{\sigma^2}{11}}\right)$

Hjälpvariabeln $\dfrac{\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)}{S \cdot \sqrt{\frac{2}{11}}} \sim t(40)$

Instängning ger

$I_{\mu_i - \mu_j} = \left(\bar{x}_i - \bar{x}_j \mp t \cdot s\sqrt{\tfrac{2}{11}}\right) = (\bar{x}_i - \bar{x}_j \mp 2.169)$

där $t = 2.42$ ges i $t(40)$-tabell.

$I_{\mu_4 - \mu_2} = (34.145 - 32.991 \mp 2.169) = (-1.015, 3.323)$

$I_{\mu_4 - \mu_3} = (34.145 - 31.418 \mp 2.169) = (0.558, 4.896)$

Tvättmedel nr 4 tycks ge vitare tvätt än nr 3 och nr 1, men inte säkert vitare än nr 2, eftersom $0 \in I_{\mu_4 - \mu_2}$, medan $I_{\mu_4 - \mu_3}$ och $I_{\mu_4 - \mu_1}$ bara innehåller positiva värden.

5a) $x = 32$ är observation av $X \sim Po(\lambda)$.

Vidare är X appr $N(\lambda, \sqrt{\lambda})$ om $\lambda > 15$.

$\hat{\lambda} = x$.

Teststorhet: $x = 32$.

H_0 förkastas om $x > c$ pga H_1.

$0.05 = P(X > c \text{ om } H_0 \text{ är sann}) =$

$= P(X > c \text{ om } \lambda = 25) \simeq 1 - \Phi\left(\frac{c-25}{\sqrt{25}}\right)$

$\Phi\left(\frac{c-25}{5}\right) = 0.95$ ger $\frac{c-25}{5} = 1.645$ och $c = 33.2$

$x = 32 < 33.2$; H_0 kan inte förkastas.

b) Under de t närmaste månaderna får man y beställningar.

Den s.v. $Y \sim Po(\lambda t)$ appr $N(\lambda t, \sqrt{\lambda t})$.

Teststorhet: y

H_0 förkastas om $y > b$.

$0.05 = P(Y > b \text{ om } \lambda = 25) \simeq 1 - \Phi\left(\frac{b-25t}{\sqrt{25t}}\right)$

$0.90 = P(Y > b \text{ om } \lambda = 35) \simeq 1 - \Phi\left(\frac{b-35t}{\sqrt{35t}}\right)$

$$\begin{cases} \Phi\left(\frac{b-25t}{\sqrt{25t}}\right) = 0.95 \\ \Phi\left(-\frac{b-35t}{\sqrt{35t}}\right) = 0.90 \end{cases} \qquad \begin{cases} \frac{b-25t}{\sqrt{25t}} = 1.645 \\ \frac{35t-b}{\sqrt{35t}} = 1.282 \end{cases}$$

$$\begin{cases} b - 25t = 8.225\sqrt{t} \\ 35t - b = 7.584\sqrt{t} \end{cases} \text{ ger } 10t = 15.809\sqrt{t} \text{ dvs}$$

$\sqrt{t} = 1.5809$ och $t \simeq 2.5$

Man ska mäta under minst 2.5 månader.

6a) $L(\lambda) = \prod_{i=1}^{n} \left[\frac{6-\lambda}{6+\lambda} \cdot 2 \cdot \frac{\lambda^{x_i}}{6^{x_i}} \right] = \frac{2^n}{6^{\Sigma x_i}} \cdot \left(\frac{6-\lambda}{6+\lambda}\right)^n \cdot \lambda^{\Sigma x_i}$

eftersom alla $x_i > 0$.

$l(\lambda) = \ln L(\lambda) = \text{konst} + n(\ln(6-\lambda) - \ln(6+\lambda)) + \Sigma x_i \ln\lambda$

$$l'(\lambda) = \frac{n}{6-\lambda}(-1) - \frac{n}{6+\lambda} + \Sigma x_i \cdot \frac{1}{\lambda} =$$

$$= -n \cdot \frac{12}{(6-\lambda)(6+\lambda)} + \frac{\Sigma x_i}{\lambda} = \frac{-12n\lambda + \Sigma x_i(36-\lambda^2)}{(6-\lambda)(6+\lambda)\lambda} =$$

$$= \frac{-n(12\lambda + \bar{x}(\lambda^2 - 36))}{(6-\lambda)(6+\lambda)\lambda} = \frac{-n\bar{x}\left(\lambda^2 + \frac{12\lambda}{\bar{x}} - 36\right)}{(6-\lambda)(6+\lambda)\lambda}$$

$l'(\lambda) = 0 \Leftrightarrow \lambda^2 + \frac{12}{\bar{x}}\lambda - 36 = 0 \Leftrightarrow \lambda = -\frac{6}{\bar{x}} \overset{+}{(-)} \sqrt{\frac{36}{\bar{x}^2} + 36}$

$\bar{x} = 7.3$ ger $\hat{\lambda} = 5.234$

λ	0	$\hat{\lambda}$	6
$l'(\lambda)$	$+$	0	$-$
$l(\lambda)$	$\nearrow$		$\searrow$

Alltså max. ML-skattn. $\hat{\lambda} = 5.234$

b) $P(X_i \geq 8) = \sum_{k=8}^{\infty} \frac{6-\lambda}{6+\lambda} \cdot 2 \cdot \left(\frac{\lambda}{6}\right)^k =$

$$= \frac{6-\lambda}{6+\lambda} \cdot 2 \cdot \left(\frac{\lambda}{6}\right)^8 \cdot \frac{1}{1-\lambda/6} = \frac{12}{6+\lambda} \cdot \left(\frac{\lambda}{6}\right)^8$$

som skattas med $\frac{12}{6+\hat{\lambda}} \cdot \left(\frac{\hat{\lambda}}{6}\right)^8 = 0.358$