TEKNISKA HÖGSKOLAN I LINKÖPING
Matematiska institutionen
Matematisk statistik
EE

Kurskod: TAMS65

Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Tentamen tisdagen den 18 augusti 2009 kl 8-12.

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 tagit fram en ny metod för att bestämma järnkoncentrationen i blodplasma. Tjugo bestämningar av järnkoncentrationen för ett kontrollserum gav följande mätvärden (enhet: μ g/100ml):

96	98	99	100	103	103	104	104	105	105
106	106	107	108	108	108	110	113	114	114

med stickprovsstandardavvikelsen $s = 5.02$.
Vi kan anta att vi har oberoende observationer från $N(\mu, \sigma)$.
För den här typen av mätningar anses det önskvärt att $\sigma < 7$. Förefaller detta villkor att vara uppfyllt? Konstruera ett lämpligt 95% konfidensintervall. (1p)

b) Man väljer mellan två sorters system för en viss typ av dataservice. Det är önskvärt att svarstiden för ett visst kommando ligger under tre sekunder. Vid 66 oberoende provningar för vardera systemet noterades om svarstiden låg över eller under tre sekunder och man fick följande resultat för de båda systemen:

	Antal fall	
System	Svarstid $> 3s$	Svarstid $< 3s$
A	14	52
B	28	38

Ger detta tillräckligt underlag för att säga att det finns skillnad mellan systemen i fråga om svarstider? Besvara frågan med hjälp av ett lämpligt konfidensintervall eller test på nivån 0.05. (2p)

Det finns inget samband mellan a) och b).

2. En ingenjör har modifierat designen för en robot som används vid svetsning. Roboten anses fungera bra om den misslyckas med högst 1% av sina svetsningar och riktigt dåligt om den misslyckas med 5% av sina svetsningar. Man ska utvärdera roboten med hjälp av 100 provsvetsningar och utgående från resultatet i dem pröva

$$H_0: p = 0.01 \quad \text{mot} \quad H_1: p > 0.01$$

där $p =$ sannolikheten att roboten misslyckas med en svetsning. Man har bestämt sig för att förkasta H_0 om minst 3 av de 100 provsvetsningarna misslyckas. Olika svetsningar misslyckas oberoende av varandra.

a) Bestäm signifikansnivån. (1.5p)

b) Beräkna styrkan för $p = 0.05$. (1.5p)

Lämplig approximation får utnyttjas både i a) och b).

3. För investerare är det mycket viktigt att förutsäga börsutvecklingen och många försök har gjorts att hitta lämpliga formler. I datamaterialet nedan finns samhörande värden den procentuella värdeförändringen de första fem dagarna i januari och värdeförändringen för hela året. Idén är att man utgående från januaridagarna ska kunna förutsäga utvecklingen under det kommande året (man brukar tala om "early warning").

Year	% Change for First 5 Days in Jan., x	% Change for Year, y	Year	% Change for First 5 Days in Jan., x	% Change for Year, y
1950	2.0	21.8	1969	−2.9	−11.4
1951	2.3	16.5	1970	0.7	0.1
1952	0.6	11.8	1971	0.0	10.8
1953	−0.9	−6.6	1972	1.4	15.6
1954	0.5	45.0	1973	1.5	−17.4
1955	−1.8	26.4	1974	−1.5	−29.7
1956	−2.1	2.6	1975	2.2	31.5
1957	−0.9	−14.3	1976	4.9	19.1
1958	2.5	38.1	1977	−2.3	−11.5
1959	0.3	8.5	1978	−4.6	1.1
1960	−0.7	−3.0	1979	2.8	12.3
1961	1.2	23.1	1980	0.9	25.8
1962	−3.4	−11.8	1981	−2.0	−9.7
1963	2.6	18.9	1982	−2.4	14.8
1964	1.3	13.0	1983	3.2	17.3
1965	0.7	9.1	1984	2.4	1.4
1966	0.8	−13.1	1985	−1.9	26.3
1967	3.1	20.1	1986	−1.6	14.6
1968	0.2	7.7			

Det är känt att olika sorters kriser har stor inverkan på börsen. I det aktuella datamaterialet finns två krisår, nämligen 1973 och 1974 då vi hade den så kallade "energikrisen". Vi låter

$Y =$ procentuell förändring under året

$x =$ procentuell förändring de första fem dagarna i januari

$$u = \begin{cases} 1 & \text{för krisåren 1973 och 1974} \\ 0 & \text{annars} \end{cases}$$

Vi har betraktat de aktuella åren som ett slumpmässigt urval av börsår och analyserat data enligt modellen

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

där $\varepsilon \sim N(0, \sigma)$, se nedan.

a) Gör x nytta som förklaringsvariabel? Motivera ditt svar med hjälp av ett lämpligt konfidensintervall eller test. Nivå 5%. (1p)

b) Konstruera ett 95% intervall som beskriver den tänkbara börsutvecklingen för ett år som inte är krisår och då börsen steg 1% under de första fem dagarna i januari. Är du nöjd med intervallet? (2p)

Minitabanalys till uppgift 3.

```
MTB > set c5
DATA> 23(0)
DATA> 1 1
DATA> 12(0)
DATA> end

MTB > Name m5 "XPXI5"
MTB > Regress 'Y' 2  'x'  'u';
SUBC>   XPXInverse 'XPXI5';
SUBC>   Constant;
SUBC>   Brief 2.
```

Regression Analysis: Y versus x, u

```
The regression equation is
Y = 9.76 + 3.34 x - 33.3 u


Predictor       Coef  SE Coef
Constant       9.759    2.240
x              3.343    1.018
u            -33.309    9.571


S = 13.1591   R-Sq = 40.9%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  4077.5  2038.7  11.77  0.000
Residual Error  34  5887.5   173.2
Total           36  9965.0
```

Data Display

```
Matrix XPXI5

 0.0289760  -0.0015559  -0.028976
-0.0015559   0.0059843   0.001556
-0.0289760   0.0015559   0.528976
```

$= (\mathbf{X}^T\mathbf{X})^{-1}$

4. I en maskin monteras ett vektyg i en hållare och man "drar åt". I ett försök utnyttjade man tre olika åtdragningsmoment mellan hållare och verktyg och i varje försök bestämde man koordinaterna för skärspetsen på verktyget. För varje åtdragningsmoment gjordes 21 mätningar. Resutat för x-koordinaterna (enhet : $10^{-4} mm$):

Moment	$\bar{x}_{i.}$	s_i
60	-20.167	4.1363
70	4.3000	3.1310
80	9.7667	4.1662

Modell: Vi har oberoende observationer från $N(\mu_i, \sigma_i)$, där $i = 1, 2, 3$.

a) Pröva på nivån 0.10 hypotesen

$$H_0 : \sigma_2 = \sigma_3 \qquad \text{mot} \qquad H_1 : \sigma_2 \neq \sigma_3 .$$

(Egentligen ska även de andra standardavvikelserna jämföras, men du behöver inte genomföra de testen.) (1p)

b) Anta nu att $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$. Gör parvisa jämförelser mellan väntevärdena genom att konstruera lämpliga konfidensintervall vart och ett med konfidensgrad 99%. Två åtdragningsmoment anses likvärdiga om $|\mu_i - \mu_k| < 10$. Är detta uppfyllt för något par? (2p)

5. En firma säljer en viss typ av enheter. Antalet efterfrågade enheter av en viss typ under en arbetsvecka är $Po(\lambda)$ och antalen för olika veckor är oberoende. Under de senaste 10 veckorna har man sålt 240 enheter vilket verkar innebära en ökning jämfört med tidigare.

a) Konstruera ett tvåsidigt konfidensintervall för λ med approximativ konfidensgrad 95%. (1.5p)

b) Det är viktigt att alltid ha enheter av den aktuella typen i lager. Man har därför haft rutinen att beställa hem nya enheter då det bara finns 16 kvar av den aktuella typen. Leveranstiden för nya enheter är exakt en arbetsvecka. Låt p vara sannolikheten att lagret blir tomt under en leveransvecka. Konstruera med hjälp av resultatet ovan ett 95% konfidensintervall för p. Lämplig approximation får utnyttjas. Bör man ändra sina beställningsrutiner? (1.5p)

6. Låt $y_1, \ldots, y_n$ vara observationer av oberoende stokastiska variabler $Y_1, \ldots, Y_n$, sådana att $Y_j = \beta x_j + \varepsilon_j$, där $\varepsilon_j \sim N(0, \sigma)$ och $x_1, \ldots, x_n$ är fixa givna tal.

Bestäm minsta-kvadrat-skattningen av β, undersök om den är väntevärdesriktig och bestäm variansen för skattningsvariabeln. (3p)

Lösningar till tentamen i TAMS65, 2009-08-18.

1a) Vi gör ett uppåt begr. konfidensintervall för σ.

Hjälpvariabeln $\frac{19S^2}{\sigma^2} \sim \chi^2(19)$ och vi får

$P(10.11 \leq \frac{19S^2}{\sigma^2}) = 0.95$

$P(\sigma^2 \leq \frac{19S^2}{10.11}) = 0.95$

vilket ger $I_\sigma = (0, s\sqrt{19/10.11}) = (0, 6.88)$

Alltså är $\sigma \leq 6.88 < 7$ med stor sannolikhet.

b) Låt p_A och p_B vara sannolikheterna för svarstider under 3s. Vi konstruerar $I_{p_A - p_B}$.

$\hat{p}_A = \frac{x}{66} = \frac{52}{66} = 0.7879$ där $X \sim Bin(66, p_A)$

$\hat{p}_B = \frac{y}{66} = \frac{38}{66} = 0.5758$ där $Y \sim Bin(66, p_B)$

Normalapproximation är OK, eftersom

$66\hat{p}_A(1-\hat{p}_A) > 10$ och $66\hat{p}_B(1-\hat{p}_B) > 10$.

Den s.v. $\hat{P}_A - \hat{P}_B$ appr $N\left(p_A - p_B, \sqrt{\frac{p_A(1-p_A)}{66} + \frac{p_B(1-p_B)}{66}}\right)$

Hjälpvariabeln $\frac{\hat{P}_A - \hat{P}_B - (p_A - p_B)}{\sqrt{\frac{\hat{P}_A(1-\hat{P}_A)}{66} + \frac{\hat{P}_B(1-\hat{P}_B)}{66}}}$ appr $N(0,1)$

och den ger

$$I_{p_A - p_B} = \left(\hat{p}_A - \hat{p}_B \mp 1.96\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{66} + \frac{\hat{p}_B(1-\hat{p}_B)}{66}}\right) =$$

$$= (0.2121 \mp 0.1547) \approx (0.057, 0.367)$$

Bara positiva värden. System A är bättre med stor sannolikhet.

<u>Alt.</u> Man kan göra homogenitetstest i stället.

2. Vi har alltså teststorheten x = antalet misslyckade provsvetsningar bland 100 st.

H_0 förkastas om $x \geq 3$.

a) Signifikansnivån: $\alpha = P(H_0$ förkastas om H_0 sann$) =$

$= P(X \geq 3$ om $p = 0.01) = 1 - P(X \leq 2$ om $p = 0.01)$

H_0 sann: $X \sim Bin(100, 0.01)$ appr $Po(1)$.

Tabell ger $\alpha \approx 1 - (0.3679 + 0.3679 + 0.1839) \approx 0.080$

b) Styrkan för $p = 0.05$: $1 - \beta = P(H_0$ förk. om $p = 0.05) =$

$= P(X \geq 3$ om $p = 0.05) = 1 - P(X \leq 2$ om $p = 0.05)$

Om $p = 0.05$ så $X \sim Bin(100, 0.05)$ appr $Po(5)$.

Tabell ger $1 - \beta \approx 1 - (0.0067 + 0.0337 + 0.0842) \approx$

≈ 0.875

3a) Vi konstruerar I_{β_1} med konfidensgrad 95%.

$I_{\beta_1} = (\hat{\beta}_1 \mp t \cdot s\sqrt{h_{11}}) = (3.343 \mp 2.032 \cdot 1.018) =$

$= (3.343 \mp 2.069) = (1.274, 5.412)$

där $t = 2.032$ ges i $t(34)$-tabell.

$0 \notin I_{\beta_1}$. Alltså gör x nytta som förklaringsvariabel.

3b) Vi konstruerar ett prediktionsintervall

för $Y_0 = \beta_0 + \beta_1 + \varepsilon_0 = (1\ 1\ 0)\beta + \varepsilon_0$ där $\varepsilon_0 \sim N(0, \sigma)$.

Predikteringsfelet $Y_0 - (1\ 1\ 0)\hat{\beta} \sim N(0, \sigma\sqrt{1+k})$

där $k = (1\ 1\ 0)(X^T X)^{-1}\begin{pmatrix}1\\1\\0\end{pmatrix} = 0.03185$

σ^2 skattas med $s^2 = Q_{RES}/34$; $s = 13.1591$;
frihetsgrad: 34.

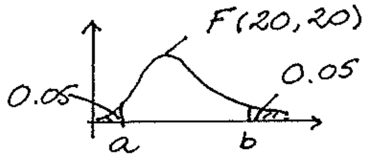
Hjälpvariabeln $\dfrac{Y_0 - (1\,1\,0)\hat{\beta}}{s\sqrt{1.03185}} \sim t(34)$ och ger

$I_{Y_0} = ((1\ 1\ 0)\hat{\beta} \mp 2.032 \cdot s\sqrt{1.03185}) =$

$= (13.102 \mp 27.162) \approx (-14, 40)$

Intervallet är väldigt långt så prognosmetoden är inte så säker, vilket man kanske kunde misstänka.

4a) Teststorhet: $V = \frac{s_3^2}{s_2^2} = 1.77$



H_0 förkastas om $v < a \iff \frac{1}{v} > \frac{1}{a}$

eller om $v > b$. Den s.v. $V \sim F(20,20)$
om H_0 är sann.
Tabell ger $b = 2.12$.
Eftersom också $\frac{1}{V} \sim F(20,20)$ får vi $\frac{1}{a} = 2.12$
d.v.s. $a = 0.47$.

$0.47 < v < 2.12$. H_0 kan inte förkastas ens på nivån
0.10. Det verkar inte orimligt att anta $\sigma_1 = \sigma_2 = \sigma_3$.

b) Vi konstruerar intervallen $I_{\mu_i - \mu_k}$
$\hat{\mu}_i - \hat{\mu}_k = \bar{x}_{i\cdot} - \bar{x}_{k\cdot}$; den s.v. $\bar{X}_{i\cdot} - \bar{X}_{k\cdot} \sim$
$$\sim N\left(\mu_i - \mu_k, \sqrt{\frac{\sigma^2}{21} + \frac{\sigma^2}{21}}\right)$$

Sammanvägd σ^2-skattning
$$s^2 = \frac{20s_1^2 + 20s_2^2 + 20s_3^2}{60} = 14.7564; \quad s = 3.8414; \quad \text{fr. grad: } 60.$$

Hjälpvariabeln $\dfrac{\bar{X}_i - \bar{X}_k - (\mu_i - \mu_k)}{S \cdot \sqrt{2/21}} \sim t(60)$ och

ger
$$I_{\mu_i - \mu_k} = \left(\bar{x}_i - \bar{x}_k \mp 2.66 \cdot s\sqrt{\tfrac{2}{21}}\right) = (\bar{x}_i - \bar{x}_k \mp 3.1534)$$

$$I_{\mu_1 - \mu_2} = (-24.4670 \mp 3.1534) \simeq (-27.6, -21.3)$$
$$I_{\mu_1 - \mu_3} = (-29.9337 \mp 3.1534) \simeq (-33.1, -26.8)$$
$$I_{\mu_2 - \mu_3} = (-5.4667 \mp 3.1534) \simeq (-8.6, -2.3)$$

$2.3 \le |\mu_2 - \mu_3| \le 8.6$. Åtdragningsmomenten 70 och 80 kan alltså anses likvärda medan momentet 60 tydligt avviker.

5a) $y = 240$ är observation av $Y \sim Po(10\lambda)$.
$$\hat{\lambda} = \frac{y}{10} = 24$$
$$E(\hat{\Lambda}) = E\left(\frac{Y}{10}\right) = \frac{1}{10}E(Y) = \frac{1}{10} \cdot 10\lambda = \lambda$$

$Var(\hat{\Lambda}) = Var(\frac{Y}{10}) = \frac{1}{100} Var(Y) = \frac{1}{100} \cdot 10\lambda = \frac{\lambda}{10}$

Eftersom $10\hat{\lambda} > 15$ gäller att Y är approximativt normalfördelad och då följer att den s.v.

$\hat{\Lambda}$ appr $N(\lambda, \sqrt{\lambda/10})$.

Hjälpvariabeln $\frac{\hat{\Lambda} - \lambda}{\sqrt{\hat{\Lambda}/10}}$ appr $N(0,1)$ och ger

$I_\lambda = (\hat{\lambda} \mp 1.96\sqrt{\hat{\lambda}/10}) = (24 \mp 3.04) = (20.96,\ 27.04)$

d.v.s. $20.96 \le \lambda \le 27.04$ med stor sannolikhet.

b) X = antalet beställningar under leveranstiden 1 vecka.

$X \sim Po(\lambda)$; $p = P(X \ge 16) = 1 - P(X \le 15)$

X appr $N(\lambda, \sqrt{\lambda})$ eftersom $\hat{\lambda} = 24 > 15$.

Alltså $p \approx 1 - \Phi(\frac{15-\lambda}{\sqrt{\lambda}}) = \Phi(\sqrt{\lambda} - \frac{15}{\sqrt{\lambda}})$

Eftersom p är en växande funktion av λ gäller

$20.96 \le \lambda \le 27.04 \Rightarrow \Phi(1.302) \le p \le \Phi(2.315)$

d.v.s. $0.904 \le p \le 0.990$

Risken för tomt lager är för stor. Man bör beställa innan lagret sjunkit så mycket som till 16.

6. Vi får $\hat{\beta}$ genom att minimera

$$Q(\beta) = \sum_{j=1}^{n} (y_j - E(Y_j))^2 = \sum_{j=1}^{n} (y_j - \beta x_j)^2$$

$$Q'(\beta) = \sum_{j=1}^{n} 2(y_j - \beta x_j)(-x_j) = -2\left(\sum_{j=1}^{n} y_j x_j - \beta \sum_{j=1}^{n} x_j^2\right)$$

$$Q'(\beta) = 0 \quad \text{för} \quad \hat{\beta} = \frac{\sum_{j=1}^{n} y_j x_j}{\sum_{j=1}^{n} x_j^2}$$

$Q''(\beta) = 2\sum_{j=1}^{n} x_j^2 > 0$; alltså minimum så $\hat{\beta}$ är MK-skattningen.

($\underline{\text{Alt.}}$ $\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \overbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}^{\mathbb{X}} \beta + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

Då är MK-skattn. (se regr. kompendiet)

$\hat{\beta} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbb{y} = \sum_{j=1}^{n} x_j y_j \Big/ \sum_1^n x_i^2 .)$

Låt $A_x = \sum_1^n x_i^2$

$$E(\hat{\beta}) = E\left(\frac{1}{A_x} \sum_{j=1}^n x_j Y_j\right) = \frac{1}{A_x} \sum_{j=1}^n x_j E(Y_j) =$$

$$= \frac{1}{A_x} \sum_{j=1}^n x_j \cdot \beta x_j = \beta \cdot \frac{\sum_1^n x_j^2}{\sum_1^n x_i^2} = \beta$$

Alltså är $\hat{\beta}$ väntevärdesriktig.

$$Var(\hat{\beta}) = Var\left(\frac{1}{A_x} \sum_{j=1}^n x_j Y_j\right) = \frac{1}{A_x^2} \sum_{j=1}^n x_j^2 \overbrace{Var(Y_j)}^{\sigma^2} =$$

$$= \sigma^2 \cdot \frac{\sum_1^n x_j^2}{\left(\sum_1^n x_i^2\right)^2} = \frac{\sigma^2}{\sum_1^n x_j^2}$$