Kurskod: TAMS65
Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Tentamen måndagen den 30 maj 2016 kl 8–12.

Hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen och/eller formelsamling "Formel- och tabellsamling i matematisk statistik TAMS65 (Martin Singull)". Inga anteckningar i formelsamlingarna är tillåtet. Miniräknare med tömda minnen.
Betygsgränser: 8-11 poäng ger betyg 3, 11.5-14.5 ger betyg 4 och 15-18 poäng ger betyg 5.
Examinator: Martin Singull, 013–281447
Resultatet meddelas *normalt* via LADOK inom 12 arbetsdagar.
Tydliga svar och motiveringar krävs till varje uppgift.

1. Man har gjort en undersökning om blodgruppsfördelningen i USA och Europa är samma. Följande resultat har erhållits

Blodgrupp	Europa	USA
A	95	125
B	50	70
0	45	90
AB	10	15

Testa på nivå 5% om blodgruppsfördelningarna är samma. (2p)

2. Man har för åtta olika lösningar mätt jonkoncentrationen vid två tidpunkter. Följande observationer erhölls

Lösning nr	1:a mättillfället	2:a mättillfället
1	10.77	10.20
2	7.90	7.42
3	11.50	11.04
4	6.74	6.10
5	9.02	8.40
6	10.71	10.04
7	5.50	4.83
8	11.88	11.40

Kan man påstå att den förväntade jonkoncentrationen i en lösning sjunker minst 0.5 enheter mellan de två mättidpunkterna? Sätt upp en lämplig modell för data och utför en hypotesprövning på signifikansnivå 5%. (3p)

3. Vid undersökning av ett smärtstillande preparat studerar man effektiviteten för olika mängder (enhet = mg) av de aktiva beståndsdelarna A, B och C. Resultat

A	B	C	Effektivitet (%)
x_1	x_2	x_3	y
15	20	10	47
15	20	20	54
15	30	10	58
15	30	20	66
30	20	10	59
30	20	20	67
30	30	10	71
30	30	20	83
45	20	10	72
45	20	20	82
45	30	10	85
45	30	20	94

Man analyserar observationerna enligt modellen

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

där $\varepsilon \sim N(0, \sigma)$ (oberoende). I analysen har vi fått den skattad regressionslinje och variansanalys:

$$y = -2.33 + 0.900x_1 + 1.27x_2 + 0.900x_3,$$

i	$\widehat{\beta_i}$	$d(\widehat{\beta_i})$
0	-2.333	2.200
1	0.90000	0.02805
2	1.26667	0.06872
3	0.90000	0.06872

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	3	2182.33
RES	8	11.33
TOT	11	2193.67

och

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 3.41667 & -0.01667 & -0.08333 & -0.05000 \\ -0.01667 & 0.00056 & -0.00000 & 0.00000 \\ -0.08333 & -0.00000 & 0.00333 & 0.00000 \\ -0.05000 & 0.00000 & 0.00000 & 0.00333 \end{pmatrix}$$

a) Är det bra anpassning mellan de observerade värdena och modellen? Beräkna och tolka R^2. (1p)

b) Pröva på nivån 1% hypotesen

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{mot} \quad H_1 : \text{minst en av } \beta_1, \beta_2, \beta_3 \neq 0$$

(1p)

c) Man överväger att tillverka en tablett med 30 mg av A, 20 mg av B och 10 mg av C. Låt Y^* och Y^{**} vara effektiviteterna (mätta på samma sätt som i undersökningen), då man tar 1 respektive 1.5 tablett. Konstruera ett konfidensintervall för $E(Y^{**}) - E(Y^*)$ med konfidensgraden 95%. (2p)

4. En stokastisk variabel X kan anta värden mellan 0 och 1. Dess täthetsfunktion antas vara

$$f(x) = \theta_1 + \theta_2 x, \quad \text{för } 0 \leq x \leq 1.$$

Fem oberoende observationer på X har givit: 0.85 0.53 0.19 0.68 0.70

   a) Bestäm det villkor som θ_1 och θ_2 måste uppfylla för att $f(x)$ ska vara en täthetsfunktion. (1p)
   b) Skatta den återstående parametern med momentmetoden. (1p)
   c) Beräkna variansen för skattningen. (1p)

5. I en poissonprocess inträffar impulser med den okända intensiteten λ impulser / tidsenhet. Under tio tidsenheter har man noterat fem impulser.

   a) Test hypotesen $H_0 : \lambda \leq 0.2$ mot $H_1 : \lambda > 0.2$ på cirka nivån 5%. (1.5p)
   b) Beräkna testets styrka för λ-värdena 0.3, 0.5 och 0.8. (1.5p)

6. Antag att $X_1, X_2, ..., X_n$ och $Y_1, Y_2, ..., Y_m$ är två oberoende stickprov på $Re(0, a)$- respektive $Re(0, 2a)$-fördelning. Finn med hjälp av stickprovsmedelvärdena $\bar{X}$ och $\bar{Y}$ en väntevärdesriktig punktskattning av parametern a med så liten varians som möjligt. (3p)

Kurskod: TAMS65
Lösningar

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningsförslag till tentamen måndagen den 30 maj 2016 kl 8–12.

1)

$$Q = \sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(N_{ij}-n_i\hat{p}_j)^2}{n_i\hat{p}_j} = 3.57 < c = \chi^2_{0.95}(3) = 7.82$$

$\underline{\underline{\text{Vi kan inte förkasta hypotesen om samma blodgruppsfördelning i Europa och USA.}}}$

2) Bilda $Z_i = X_i - Y_i \sim N(\Delta, \sigma)$ vilket ger $\bar{Z} \sim N\left(\Delta, \frac{\sigma}{\sqrt{n}}\right)$. Vi vill testa hypotesen

$$H_0 : \Delta \leq 0.5 \quad \text{mot} \quad H_1 : \Delta > 0.5 \quad \text{nivå } 5\%.$$

Teststorhet

$$T = \frac{\bar{Z} - 0.5}{S/\sqrt{8}} \sim t(7) \quad \text{under } H_0.$$

Observationerna ger att $t = 2.34$. Förkasta H_0 om $t > c = t_{0.95}(7) = 1.89$. Alltså, $\underline{\underline{\text{förkasta } H_0\text{, jonkoncentrationen verkar sjunka mer än 0.5 enheter.}}}$

3a) $R^2 = \dfrac{SS_{REGR}}{SS_{TOT}} = \underline{\underline{99.5\%}}$, det verkar vara väldigt bra anpassning dvs. den givna modellen verkar förklara variationen i datan väldigt bra.

b) Teststorhet $v = \dfrac{SS_{REGR}/3}{SS_{RES}/8} = 513.49$. Under H_0 gäller att $V \sim F(3, 8)$. Förkasta H_0 om $v > c = F_{0.99}(3, 8) = 7.59$. Alltså, $\underline{\underline{\text{förkasta } H_0}}$ och det finns något i modellen som är användbart.

c) $E(Y^{**}) - E(Y^{*}) = 15\beta_1 + 10\beta_2 + 5\beta_3 = \boldsymbol{u}'\boldsymbol{\beta}$ där $\boldsymbol{u} = (0, 15, 10, 5)'$ och $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$.

Punktskattning: $\boldsymbol{u}'\widehat{\boldsymbol{\beta}} = 30.6667$ som är en observation från en normalfördelning med $\mathrm{E}(\boldsymbol{u}'\widehat{\boldsymbol{\beta}}) = \boldsymbol{u}'\boldsymbol{\beta}$ och $\mathrm{var}(\boldsymbol{u}'\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{u}'(\boldsymbol{X}\boldsymbol{X}')^{-1}\boldsymbol{u} = 0.5423\sigma^2$ där variansen σ^2 skattas med $s^2 = \dfrac{SS_{RES}}{8} = 1.4167$ vilket ger $s = 1.190$.

Konfidensintervallet ges nu av

$$I_{E(Y^{**})-E(Y^{*})} = \left(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} \mp t_{0.975}(8)s\sqrt{\boldsymbol{u}'(\boldsymbol{X}\boldsymbol{X}')^{-1}\boldsymbol{u}}\right) = \underline{\underline{(28.6; 32.7)}}.$$

4a) $1 = \int_0^1 f(x)dx = \theta_1 + \dfrac{\theta_2}{2}$ ger $\underline{\underline{\theta_2 = 2(1-\theta_1)}}$, dvs. $f(x) = \theta + 2(1-\theta)x$, $0 < x < 1$.
Det går även bra att lösa ut θ_1 istället.

b) $\mathrm{E}(X) = \int_0^1 xf(x)dx = \dfrac{\theta}{2} + \dfrac{2}{3}(1-\theta) = \dfrac{2}{3} - \dfrac{\theta}{6}$

Momentmetoden ger $\dfrac{2}{3} - \dfrac{\theta}{6} = \bar{x}$ och $\underline{\underline{\hat{\theta} = 4 - 6\bar{x} = 0.46}}$

c) $\mathrm{E}(X^2) = \int_0^1 x^2 f(x)dx = \dfrac{\theta}{3} + \dfrac{1-\theta}{2}$

$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \dfrac{\theta}{3} + \dfrac{1-\theta}{2} - \left(\dfrac{2}{3} - \dfrac{\theta}{6}\right)^2 = \dfrac{2 + 2\theta - \theta^2}{36}$ vilket ger

$\mathrm{var}(\hat{\theta}) = \mathrm{var}(4 - 6\bar{X}) = 36\dfrac{\mathrm{var}(X)}{5} = \underline{\underline{\dfrac{2 + 2\theta - \theta^2}{5}}}$

5a) Teststorhet X = ”antalet impulser under 10 tidsenheter” $\sim Po(10\lambda)$. Förkasta H_0 om $X \geq c$. Under H_0 gäller att $X \sim Po(2)$. Vi har alltså

$$\begin{aligned} 5\% &\approx P(\text{förkasta } H_0 \text{ då } H_0 \text{ är sann}) \\ &= P(X \geq c \text{ då } X \sim Po(2)) \end{aligned}$$

som ger att $c = 5$. $\underline{\underline{\text{Eftersom } x = 5 \geq 5 = c \text{ så kan vi förkasta } H_0}}$ på nivån ca 5% ($\alpha = 5.26\%$).

b)

$$\begin{aligned} h(0.3) &= P(X \geq 5 \text{ då } X \sim Po(3)) = \underline{\underline{18.5\%}}, \\ h(0.5) &= P(X \geq 5 \text{ då } X \sim Po(5)) = \underline{\underline{55.9\%}}, \\ h(0.8) &= P(X \geq 5 \text{ då } X \sim Po(8)) = 1 - P(X \leq 4 \text{ då } X \sim Po(8)) = \underline{\underline{90.0\%}}. \end{aligned}$$

6) $\hat{a} = \alpha\bar{X} + \beta\bar{Y}$

$\mathrm{E}(\hat{a}) = \alpha\,\mathrm{E}(\bar{X}) + \beta\,\mathrm{E}(\bar{Y}) = \alpha\dfrac{a}{2} + \beta\dfrac{2a}{2} = a\left(\dfrac{\alpha}{2} + \beta\right) = a$ ger att $\beta = 1 - \dfrac{\alpha}{2}$

$$\begin{aligned} \mathrm{var}(\hat{a}) &= \alpha^2\,\mathrm{var}(\bar{X}) + \left(1 - \frac{\alpha}{2}\right)^2 \mathrm{var}(\bar{Y}) = \alpha^2\,\mathrm{var}(\bar{X}) + \left(1 - \frac{\alpha}{2}\right)^2 \mathrm{var}(\bar{Y}) \\ &= \alpha^2 \frac{a^2}{12n} + \left(1 - \frac{\alpha}{2}\right)^2 \frac{4a^2}{12m} = \frac{a^2}{12}\left(\frac{\alpha^2}{n} + \left(1 - \frac{\alpha}{2}\right)^2 \frac{4}{m}\right) \end{aligned}$$

$$\frac{\partial\,\mathrm{var}(\hat{a})}{\partial\alpha} = \frac{a^2}{12}\left(\frac{2\alpha}{n} - \frac{2}{2}\left(1 - \frac{\alpha}{2}\right)\frac{4}{m}\right) = 0$$

ger $\alpha = \dfrac{2n}{m+n}$

Alltså, $\underline{\underline{\hat{a} = \dfrac{2n}{m+n}\bar{X} + \dfrac{m}{m+n}\bar{Y}}}$ är en vvr skattning av a och har minimal varians bland linjärkombinationerna av $\bar{X}$ och $\bar{Y}$.