Kurskod: TAMS65

Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Tentamen lördagen den 21 augusti 2010 kl 8-12.

Hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen samt miniräknare med tömda minnen. Inga anteckningar i formelsamlingen är tillåtet.
Betygsgränser: 8-11 poäng ger betyg 3, 11.5-14.5 ger betyg 4 och 15-18 poäng ger betyg 5.
Examinator: Martin Ohlson
Resultatet meddelas normalt via LADOK inom 12 arbetsdagar.
Tydliga svar och motiveringar krävs till varje uppgift.

1. På ett kasino kan man spela olika tärningsspel där det är viktigt hur högt man slår. Givitvis är det därför av intresse om tärningen som används är en välbalanserad tärningen eller inte. För att undersöka detta slogs 240 kast med följande utfall.

	Antal, N_i
Ettor	55
Tvåor	41
Treor	34
Fyror	37
Femmor	27
Sexor	46

Undersök med lämpligt χ^2-test (nivå 0.05) om det är en välbalanserad tärning, dvs. om $p_i = 1/6$, för alla $i = 1, \ldots, 6$ eller om minst ett $p_i \neq 1/6$ för något $i = 1, \ldots, 6$. (2p)

2. I uppsatsen *"Effect of Refrigeration on the Potassium Bitartrate Stability and Composition of Italian Wines"* (A. Versari et al., Italian Journal of Food Science 2002:45-52) diskuteras mängden vinsyra i åtta olika viner som har utsatts för en kylstabiliseringsprocess. Nedan finns mätningar av mängden vinsyran (enhet g/L) i de åtta vinerna före och efter kylbehandlingen.

Vin	Före	Efter
1	2.86	2.59
2	2.85	2.47
3	1.84	1.58
4	1.60	1.56
5	0.80	0.78
6	0.89	0.66
7	2.03	1.87
8	1.90	1.71

Sänker kylbehandlingen vinsyran i vinet? Motivera ditt svar med hjälp av ett lämpligt 95% konfidensintervall. Normalfördelning kan förutsättas. (3p)

3. En snabbmatskedja vill undersöka hur dess försäljning påverkas av restaurangens läge, dvs. hur tätbebyggt område den ligger i och om den ligger längs en motorväg, i ett köpcentrum eller på gatan inne i en stad. Följande data visar försäljningen för 16 restauranger.

Antal hushåll i närområdet x_1 (i tusental)	Läge	Försäljning y (i tusentals dollar)
155	Motorväg	135.27
93	Motorväg	72.74
128	Motorväg	114.95
114	Motorväg	102.93
158	Motorväg	131.77
183	Motorväg	160.91
178	Köpcentrum	179.86
215	Köpcentrum	220.14
172	Köpcentrum	179.64
197	Köpcentrum	185.92
207	Köpcentrum	207.82
95	Köpcentrum	113.51
224	"Gata i stad"	203.98
199	"Gata i stad"	174.48
240	"Gata i stad"	220.43
100	"Gata i stad"	93.19

Data har analyserats enligt följande modell

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

där Y är försäljningen (i tusentals dollar), x_1 är antalet hushåll (i tusental) i restaurangens närområde,

$$x_2 = \begin{cases} 1 & \text{för Köpcentrum} \\ 0 & \text{för övrigt} \end{cases}, \quad x_3 = \begin{cases} 1 & \text{för "Gata i stad"} \\ 0 & \text{för övrigt} \end{cases}$$

och ε-variablerna är oberoende och $N(0, \sigma)$-fördelade. Datautskrifter från Minitab finns nedan.

a) Är modellen användbar? Testa på nivån 0.05

$$H_0: \ \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{mot} \quad H_1: \ \text{minst en } \beta_i \neq 0, \ i = 1, 2, 3.$$

(1p)

Regression Analysis: y versus x1; x2; x3

The regression equation is
y = - 1,82 + 0,878 x1 + 27,3 x2 + 7,39 x3

Predictor	Coef	SE Coef	T
Constant	-1,817	5,453	-0,33
x1	0,87782	0,03546	24,75
x2	27,298	3,620	7,54
x3	7,392	4,177	1,77

S = 5,79945   R-Sq = 98,8%   R-Sq(adj) = 98,5%

Analysis of Variance

Source	DF	SS	MS
Regression	3	33439	11146
Residual Error	12	404	34
Total	15	33842	

b) Punktskatta försäljningsskillnaden mellan två restauranger som ligger i lika tätbebyggda områden men där den ena ligger i ett köpcentrum och den andra på gatan i en stad. (1p)

c) Konstruera ett 95% konfidensintervall för $\mathrm{E}(Y)$ då antalet hushåll i närområdet är 100 (dvs. egentligen 100 tusen) och då restaurangen ligger längs en motorväg.

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.883986 & -0.0051792 & 0.034459 & 0.103947 \\ -0.005179 & 0.0000374 & -0.001452 & -0.001954 \\ 0.034459 & -0.0014522 & 0.389726 & 0.242543 \\ 0.103947 & -0.0019539 & 0.242543 & 0.518757 \end{pmatrix}.$$

(2p)

4. Under december månad år 2007 genomförde Synovate en undersökning bland allmänheten angående utbyggnad av ny kärnkraft. Målgrupp för undersökningen var den svenska allmänheten 16 år och äldre. Sammanlagt genomfördes 1037 intervjuer. Av dessa intervjuer svarade 539 personer att de var positiva till att bygga nya reaktorer. Låt p vara andelen av det svenska folket som är positiva till att bygga nya reaktorer i Sverige.

   a) Kan vi säga att majoriteten av svenska folket är positiva till kärnkraft, dvs. pröva $H_0: \ p = 0.5$ mot $H_1: \ p > 0.5$ på nivån högst 0.05. (2p)

   b) Beräkna testets styrka för $p = 0.55$. (1p)

5. Låt $x_1, \ldots, x_n$ vara n oberoende observationer från $N(a\mu, 1)$ och låt $y_1, \ldots, y_m$ vara m oberoende observationer från $N(b\mu, 1)$, där a och b är kända konstanter.

   a) Bestäm maximum-likelihood-skattningen av μ. (2p)

   b) Visa att maximum-likelihood-skattningen av μ är väntevärdesriktig. (1p)

6. Låt $y_1, \ldots, y_n$ vara oberoende observationer från de s.v. $Y_1, \ldots, Y_n$, med

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \tag{1}$$

där $\varepsilon_i \sim N(0, \sigma)$ och $x_1, \ldots, x_n$ är fixa tal.

Vi vet att maximum-likelihood-skattningen (och MK-skattningen) av β_0 och β_1 kan skrivas som

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Om vi omparametiserar modellen (1) ovan så får vi

$$Y_i = \alpha_0 + \alpha_1 (x_i - \bar{x}) + \varepsilon_i.$$

Låt $\hat{\alpha}_0$ och $\hat{\alpha}_1$ vara maximum-likelihood-skattningarna av α_0 och α_1.

a) Visa att $\hat{\alpha}_0 = \bar{y}$ (dvs. $\hat{\alpha}_0 \neq \hat{\beta}_0$) och att $\hat{\alpha}_1 = \hat{\beta}_1$. (1p)

b) Härled fördelningen för den s.v. $\hat{\alpha}_0 = \bar{Y}$. (1p)

c) Visa att $\hat{\alpha}_0$ och $\hat{\alpha}_1$ är oberoende. (1p)