Kurskod: TAMS65
Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Tentamen måndagen den 17 oktober 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. Vid en studie av ungdomars bruk av marijuana klassificerades 445 ungdomar efter såväl eget marijuanabruk som föräldrarnas bruk av alkohol och narkotika.

		Ungdomarnas marijuananbruk		
		Aldrig	Ibland	Regelbundet
Föräldras				
bruk av	Ingen	141	54	40
alkohol och	En	68	44	51
narkotika	Båda	17	11	19

Avgör med lämpligt test om föräldrars användning av alkohol och/eller narkotika har något samband med ungdomarnas bruk av marijuana. (2p)

2. En firma studerar åtta veckors försäljningssiffror i fyra stora städer. Man bedömer att försäljningen dividerat med antalet invånare är relevant jämförelsetal. Data (försäljning per invånare avrundat till heltal)

		Vecka									
		1	2	3	4	5	6	7	8	$\bar{x}$	s
Stad	1	19	14	12	17	21	16	15	18	16.50	2.88
	2	8	12	9	9	8	10	11	8	9.38	1.51
	3	13	6	6	8	11	8	9	9	8.75	2.38
	4	17	18	17	22	16	18	14	21	17.88	2.59

Antag att data är oberoende och att data i rad i är $N(\mu_i, \sigma)$, alltså med samma σ för de olika städerna. Gör tvåsidiga konfidensintervall för de olika differenserna $\mu_i - \mu_j$. Den simultana konfidensgraden ska vara minst 94%. Finns det skillnader mellan städerna? (3p)

3. För ett antal kolkraftverk med likartad reningsteknik har man mätt y = svaveldioxidutsläpp (ppm) samt x=effekt (gigawatt)

y	105	108	$\cdots$	241
x	0.419	0.449	$\cdots$	0.715

En analys enligt modellen

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon,$$

där $\varepsilon \sim N(0, \sigma)$ (oberoende), gav resultatet

$$y = 204 - 638x + 959x^2,$$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	204.46	82.95
1	-638.0	298.5
2	959.2	263.6

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	15049.3
RES	6	254.3
TOT	8	15303.6

och

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 162.38 & -581.90 & 508.70 \\ -581.90 & 2102.17 & -1850.58 \\ 508.70 & -1850.58 & 1639.75 \end{pmatrix}$$

a) Hur många kolkraftverk har man undersökt? (1p)

b) Är andragradstermen i modellen nödvändig? Genomför lämpligt test eller konfidensintervall på nivån 5%. (1p)

c) Beräkna ett 95% konfidensintervall för $E(Y)$ då kraftverkets effekt är på 0.5 gigawatt. (2p)

4. I kvalitetsarbetet inom tillverkningsindustrin genomförs en mängd mätningar med vars hjälp man kan avgöra om olika kvalitetskrav är uppfyllda. För en viss kvalitetsvariabel samlar man in under en dag 25 oberoende mätvärden $x_1, ..., x_{25}$. Vidare är det rimligt att de stokastiska variablerna $X_1, ..., X_{25}$ är oberoende och $X_i \sim N(\mu, 1.2)$, där målvärdet är $\mu = 30$. Man prövar på nivån 5% hypotesen

$$H_0 : \mu = 30 \quad \text{mot} \quad H_1 : \mu > 30.$$

a) Man har fått $\bar{x} = 30.35$. Genomför hypotesprövningen (1.5p)

b) För vilka värden på μ är testets styrka minst 75%? (1.5p)

5. Livslängden hos en viss typ av elektroniska komponenter anses ha täthetsfunktionen

$$f(x) = a^2 x e^{-ax}, \quad x \geq 0,$$

där a är en okänd positiv konstant.

Man har observerat den sammanlagda livsländen hos 50 sådana komponenter till 250.

a) Beräkna en punktskattning av a. (2p)

b) Beräkna ett 95% nedåt begränsat konfidensintervall för den förväntade livslängden hos komponenten. (2p)

6. Antalet registrerade partiklar från ett radioaktivt prov beskrivs av en Poissonprocess med intensiteten λ. Detta innebär att antalet registrerade partiklar under t sekunder är $Po(\lambda t)$ och att registreringarna under olika tidsintervall är oberoende. Mätningarna genomfördes under 100 olika tidsintervall av längden 10 sekunder och det visade sig att för 52 tidsintervall saknades registreringar helt. Punktskatta λ med hjälp av enbart denna information. (2p)

Tips: Utnyttja sannolikheten $P(X = 0)$, där X är antalet registrerade partiklar under 10 sekunder.

Kurskod: TAMS65
Lösningar

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningsförslag till tentamen måndagen den 17 oktober 2016 kl 8–12.

1) H_0 : Föräldras bruk av alkohol och/eller narkotika är oberoende ungdomarnas bruk av marijuana, mot H_1 : De är inte oberoende.

$T = \dfrac{(141-119.35)^2}{119.35} + \dfrac{(54-57.56)^2}{57.56} + ... + \dfrac{(19-11.62)^2}{11.62} = 22.37$

Nivå 1% ger $T = 22.37 > \chi^2_{0.99}(4) = 13.28 \quad \Rightarrow \quad H_0$ förkastas.

2) σ^2 skattas med $s^2 = \dfrac{7s_1^2 + 7s_2^2 + 7s_3^2 + 7s_4^2}{28} = 5.74$

Intervallen ges av $I_{\mu_i - \mu_j} = \left(\bar{x}_i - \bar{x}_j \mp t_{0.995}(28)\ s\ \sqrt{\dfrac{1}{8} + \dfrac{1}{8}} \right)$, där $t_{0.995}(28) = 2.76$, dvs.

$I_{\mu_1-\mu_2} = (7.12 \mp 3.31) \quad \Rightarrow \quad \mu_1 > \mu_2$
$I_{\mu_1-\mu_3} = (7.75 \mp 3.31) \quad \Rightarrow \quad \mu_1 > \mu_3$
$I_{\mu_1-\mu_4} = (-1.38 \mp 3.31) \quad \Rightarrow$ ingen slutsats
$I_{\mu_2-\mu_3} = (0.63 \mp 3.31) \quad \Rightarrow$ ingen slutsats
$I_{\mu_2-\mu_4} = (-8.50 \mp 3.31) \quad \Rightarrow \quad \mu_2 < \mu_4$
$I_{\mu_3-\mu_4} = (-9.13 \mp 3.31) \quad \Rightarrow \quad \mu_3 < \mu_4$

med en simultan konfidensgrad på minst 94%. Alltså, försäljningen per invånare i stad 1 och 4 är större än i 2 och 3. Men vi kan inget säga om relationen mellan 1 och 4.

3a) 9

b) $H_0 : \beta_2 = 0$ mot $H_1 : \beta_2 \neq 0$ på nivån 5%.

Teststorhet $t = \dfrac{\widehat{\beta}_2}{d(\widehat{\beta}_2)} = 3.64 > t_{0.975}(6) = 2.45 \quad \Rightarrow \quad$ Förkasta H_0, dvs. andragradstermen gör nytta.

c) Låt $\boldsymbol{u} = \begin{pmatrix} 1 & 1/2 & 1/4 \end{pmatrix}'$. Konfidensintervallet ges av

$$I_{E(Y)} = \left(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} \mp t_{0.975}(6)\ s\ \sqrt{\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}} \right) = (117.9\ ;\ 132.6),$$

där $t_{0.975}(6) = 2.45$ och $s = \sqrt{\dfrac{SS_{RES}}{6}} = 6.51$.

4a) $\widehat{\mu} = \bar{x}$ som är en observation från $\bar{X} \sim N\left(\mu, \frac{1.2}{\sqrt{25}}\right) = N(\mu, 0.24)$

Teststorhet $u = \frac{\bar{x} - 30}{0.24} = 1.458$. Förkasta H_0 om $u > c = z_{0.95} = 1.645$. Alltså, förkasta inte H_0. Vi kan inte påvisa avvikelse från målvärdet.

b) Styrkefunktionen ges av

$$\begin{aligned} h(\mu) &= P(H_0 \text{ förkastas om } \mu \text{ är det sanna värdet}) = \\ &= P\left(\frac{\bar{X} - 30}{0.24} > 1.645 \text{ om } \mu \text{ är det sanna värdet}\right) = \\ &= P\left(\bar{X} > 30.395 \text{ om } \bar{X} \sim N(\mu, 0.24)\right) = \\ &= 1 - \Phi\left(\frac{30.395 - \mu}{0.24}\right) = \Phi\left(\frac{\mu - 30.395}{0.24}\right) \end{aligned}$$

Vidare gäller att $h(\mu) > 0.75$ om $\frac{\mu - 30.395}{0.24} \geq 0.6745$ dvs. om $\mu \geq 30.557$.

5a) $E(X) = \int_0^\infty x a^2 x e^{-ax} dx = \frac{2}{a}$.

Momentmetoden ger $\frac{2}{\widehat{a}} = \bar{x} \quad \Rightarrow \quad \widehat{a} = \frac{2n}{\sum_{i=1}^{50} x_i} = \frac{100}{250} = 0.4$ (ML-metoden ger samma skattning).

b) Det gäller att $\sum_{i=1}^{50} X_i \approx N(50\mu, \sigma\sqrt{50})$ enligt CGS.

Vidare är $\sigma^2 = \text{var}(X) = E(X^2) - (E(X))^2 = \int_0^\infty x^2 a^2 x e^{-ax} dx - \left(\frac{2}{a}\right)^2 = \frac{2}{a^2}$

Vi har alltså $\sum_{i=1}^{50} X_i \approx N\left(\frac{100}{a}, \frac{10}{a}\right)$ som ger hjälpvariabeln

$$\frac{\sum_{i=1}^{50} X_i - \frac{100}{a}}{\frac{10}{a}} \approx N(0, 1).$$

Eftersom $\mu = \frac{2}{a}$ så bildar vi först ett uppåt begränsat intervall för a.

$P\left(\frac{\sum_{i=1}^{50} X_i - \frac{100}{a}}{\frac{10}{a}} \leq 1.645\right) = 0.95$ ger intervallet $I_a = (0\ ;\ 0.4658)$ som i sin tur ger intervallet för väntevärdet $I_\mu = (4.29\ ;\ \infty)$.

6) Låt X vara antalet registreringar under ett 10s-intervall. Då gäller att $X \sim Po(10\lambda)$ och $P(X = 0) = e^{-10\lambda}$.

$z = 52$ är en observation av $Z \sim Bin(100, p)$ där $p = e^{-10\lambda}$.

$\widehat{p} = 0.52$ ger $e^{-10\widehat{\lambda}} = 0.52$ dvs. $-10\widehat{\lambda} = \ln 0.52$ eller $\widehat{\lambda} = 0.0654$.