Kurskod: TAMS65
Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Tentamen tisdagen den 27 maj 2014 kl 14–18.

Hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen och/eller formelsamling ”Formel- och tabellsamling i matematisk statistik TAMS65 - vt 2014 (Martin Singull)” (formelsamlingen ska ha en vattenmarkering TAMS65 - vt 2014 i bakgrunden på varje sida). 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. Ur ett stickprov omfattande $n = 25$ observationer från $N(\mu, \sigma)$-fördelning har man beräknat stickprovsmedelvärdet $\bar{x} = 1.1070$ och stickprovsvariansen $s^2 = 2.6302$. Beräkna ett uppåt begränsat 95% konfidensintervall för μ. (2p)

2. En fabrikant vill jämföra tre olika sorters lim och gör därför likvärdiga försök, där hållfastheten hos en viss typ av limfogar bestäms. Resultat:

Sort	Observationer								Medelvärde	Stickprovs-standard-avvikelse
1	2.91	3.02	2.97	2.83	2.95	2.88	3.07	2.88	2.9388	0.0797
2	2.78	2.90	2.97	2.95	2.82	2.99	2.88	2.97	2.9075	0.0767
3	3.04	3.29	3.32	3.21	3.30	3.44	3.33	3.47	3.3000	0.1340

   Det är inte självklart att stickproven har samma standardavvikelse men vi antar ändå modellen att sort nr i ger observation x_{ij}, $j = 1, ..., 8$, där $X_{ij} \sim N(\mu_i, \sigma)$ och där de olika stokastiska variablerna X_{ij} är oberoende.

   a) Gör parvisa jämförelser mellan väntevärdena med hjälp av lämpliga konfidensintervall vart och ett med konfidensgraden 99%. (2p)

   b) Konstruera ett 99% uppåt begränsat konfidensintervall för variansen σ^2. (1p)

3. Låt $z_1, ..., z_n$ vara oberoende observation från en normalfördelad stokastisk variabel med väntevärde noll och standardavvikelse $\sqrt{2}\sigma$. Härled maximum-likelihood-skattningen av σ^2 och undersök om den är väntevärdesriktig. (3p)

4. Vid ett avloppsreningsverk i laboratorieskala genomfördes en serie experiment för att fastställa hur fosforreduktionen (y) mätt i procent beror av avloppsvattnets pH-värde (x). Följande resultat erhölls:

observation	pH-värde (x)	Fosforreduktion (y)
1	9.2	86.5
2	9.9	93.0
$\vdots$	$\vdots$	$\vdots$
20	10.0	93.6

Data har analyserats enligt modellen

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

där ε-variablerna är oberoende och normalfördelade med väntevärde noll och standardavvikelse σ.

Skattad regressionslinje: $y = -495 + 115x - 5.61x^2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	–494.65	82.84
1	114.54	15.43
2	–5.6061	0.7128

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	?	1526.65
RES	?	172.03
TOT	19	1698.68

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 678.1636 & -126.2110 & 5.8116 \\ -126.2110 & 23.5343 & -1.0858 \\ 5.8116 & -1.0858 & 0.0502 \end{pmatrix}$$

a) Hur många frihetsgrader har kvadratsummorna REGR och RES? (1p)

b) Testa på nivån 1% om andragradstermen behövs i modellen. (1p)

c) Vilken fosforreduktion kan man förvänta sig i genomsnitt då avloppsvattnet har pH-värde 9.6. Konstruera ett lämpligt 95% intervall. (2p)

5. I en låda finns två typer av radioaktivapreparat. De båda typerna emitterar i genomsnitt 3 respektive 4 partiklar per sekund. Någon har tagit bort etiketterna på preparaten, så de ser exakt likadana ut. En person tar ett preparat. Innan hon använder det vill hon undersöka vilken typ det är. Med hjälp av en räknare finner hon att 107 partiklar emigreras under 30 sekunder. Partikelemissionen förutsätts ske enligt en poissonprocess med intensiteten λ per sekund.

   a) Pröva hypotesen $H_0 : \lambda = 3$ mot $H_1 : \lambda = 4$ med ett test som approximativt har signifikansnivån 5%. (1.5p)

   b) Beräkna testets styrka för $\lambda = 4$ och tolka resultatet. (1.5p)

6. Antag att $X \sim Bin(n, p)$. Sannolikheten p skattas med $\hat{p}$ som är en observation från den stokastiska variabeln $\widehat{P} = \dfrac{X}{n}$. Vidare gäller att $\mathrm{E}(\widehat{P}) = p$ och $\mathrm{var}(\widehat{P}) = \dfrac{p(1-p)}{n}$.

   Som en skattning till $\mathrm{var}(\widehat{P})$ används ofta $\dfrac{\widehat{P}(1-\widehat{P})}{n}$. Härled väntevärdet för detta uttryck och ange därefter en väntevärdesriktig skattning av $\mathrm{var}(\widehat{P})$. (3p)

Kurskod: TAMS65
Lösningar

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningsförslag till tentamen tisdagen den 27 maj 2014 kl 14–18.

1) Hjälpvariabel $\dfrac{\bar{X}-\mu}{S/\sqrt{25}} \sim t(24)$. Stäng in, lös ut μ och ersätt med data ger intervallet

$$I_\mu = \Big(-\infty, \bar{x} + \underbrace{t_{0.95}(24)}_{=1.71}\frac{s}{5}\Big) = \underline{\underline{(-\infty, 1.66)}}$$

2a) $I_{\mu_i-\mu_j} = \left(\bar{x}_i - \bar{x}_j \mp \underbrace{t_{0.995}(21)}_{=2.83} s\sqrt{\dfrac{1}{n_i}+\dfrac{1}{n_j}}\right)$ där $s^2 = \dfrac{7s_1^2+7s_2^2+7s_3^2}{21} = 0.0101 \Rightarrow$
$s = 0.1005$ med $df = 21$ frihetsgrader och $n_1 = n_2 = n_3 = 8$.

$I_{\mu_i-\mu_j} = \left(\bar{x}_i - \bar{x}_j \mp 2.83 \cdot 0.1005\sqrt{\dfrac{1}{8}+\dfrac{1}{8}}\right) = (\bar{x}_i - \bar{x}_j \mp 0.1422)$

$I_{\mu_1-\mu_2} = (0.0313 \mp 0.1422)$ ingen skillnad
$I_{\mu_1-\mu_3} = (-0.3612 \mp 0.1422)$ $\mu_3 > \mu_1$
$I_{\mu_2-\mu_3} = (-0.3925 \mp 0.1422)$ $\mu_3 > \mu_2$
<u>Lim nr 1 och lim nr 2 är likvärdiga, medan nr 3 är bättra än både 1 och 2.</u>

b) Hjälpvariabel $\dfrac{21S^2}{\sigma^2} \sim \chi^2(21)$. Stäng in, lös ut μ och ersätt med data ger intervallet

$$I_{\sigma^2} = \left(0, \frac{21s^2}{c}\right) = \left(0, \frac{0.2121}{8.89}\right) = \underline{\underline{(0, 0.0239)}},$$

där $c = \chi^2_{0.01}(21) = 8.89$.

3a) $z_1, ..., z_n$ oberoende observationer från $Z_i \sim N(0, \sqrt{2}\sigma)$. Likelihood-funktionen är

$$L(\sigma^2) = \left(\frac{1}{\sqrt{2\pi 2\sigma^2}}\right)^n \prod_{i=1}^n e^{-\frac{z_i^2}{4\sigma^2}} = (4\pi\sigma^2)^{-n/2} e^{-\frac{1}{4\sigma^2}\sum_{i=1}^n z_i^2}$$

$$l(\sigma^2) = \ln L(\sigma^2) = -\frac{n}{2}\ln(4\pi\sigma^2) - \frac{1}{4\sigma^2}\sum_{i=1}^n z_i^2 = konst. - \frac{n}{2}\ln(\sigma^2) - \frac{1}{4\sigma^2}\sum_{i=1}^n z_i^2$$

$$\frac{\partial l}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{4(\sigma^2)^2}\sum_{i=1}^n z_i^2$$

$$\frac{\partial l}{\partial\sigma^2} = 0 \quad \Rightarrow \quad \underline{\underline{\hat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^n z_i^2}} \quad \text{max eftersom} \quad \frac{\partial^2 l}{\partial(\sigma^2)^2}\Big|_{\sigma^2=\hat{\sigma}^2} = -\frac{n}{2}\frac{1}{(\hat{\sigma}^2)^2} < 0.$$

b) Använd att $\mathrm{E}(Z_i^2) = \mathrm{var}(Z_i) + (\mathrm{E}(Z_i))^2 = \mathrm{var}(Z_i) = 2\sigma^2$ eftersom $\mathrm{E}(Z_i) = 0$. Vi har alltså

$$\mathrm{E}(\hat{\sigma}^2) = \mathrm{E}\left(\frac{1}{2n}\sum_{i=1}^{n} z_i^2\right) = \frac{1}{2n}\sum_{i=1}^{n}\mathrm{E}\left(z_i^2\right) = \frac{1}{2n}\sum_{i=1}^{n} 2\sigma^2 = \sigma^2,$$

det vill säga $\hat{\sigma}^2$ är en väntervärdesriktig skattning av σ^2.

4a) $df_{REGR} = k = 2$ och $df_{RES} = n - k - 1 = 17$

b) Bilda konfidensintervall för parametern β_2.

$$I_{\beta_2} = \left(\widehat{\beta}_2 \mp \underbrace{t_{0.995}(17)}_{=2.90} d(\widehat{\beta}_2)\right) = (-5.6061 \mp 2.90 \cdot 0.7128) = \underline{\underline{(-7.6732, -3.5390)}}$$

Det verkar som om $\beta_2 \neq 0$ och alltså gör andragradstermen nytta i modellen.

c) Bilda konfidensintevall för $\boldsymbol{u}' = \begin{pmatrix}1 & 9.6 & 92.16\end{pmatrix}'$.

Hjälpvariabeln $\dfrac{\boldsymbol{u}'\widehat{\boldsymbol{\beta}} - \boldsymbol{u}'\boldsymbol{\beta}}{s\sqrt{\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}}} \sim t(17)$ ger intervallet

$$I_{\boldsymbol{u}'\boldsymbol{\beta}} = \left(\underbrace{\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}_{=88.2758} \mp \underbrace{t_{0.975}(17)}_{=2.11} s\sqrt{\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}}\right) = \underline{\underline{(86.1, 90.5)}},$$

där variansen σ^2 skattas med $s^2 = \dfrac{172.03}{17} = 10.1194$ och $\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u} = 0.1069$.

(Om man har rätt uttryck för konfidensintervallet $I_{\boldsymbol{u}'\boldsymbol{\beta}}$ så får man 2p)

5a) $H_0 : \lambda = 3$ mot $H_1 : \lambda = 4$

Vi har att $X \sim Po(30\lambda) \approx N(30\lambda, \sqrt{30\lambda})$. Vi skattar λ med $\hat{\lambda}$ som är en observation från den s.v. $\hat{\lambda} = \dfrac{X}{30} \approx N\left(\lambda, \sqrt{\dfrac{\lambda}{30}}\right)$.

När H_0 är sann gäller att $Z = \dfrac{\hat{\lambda} - 3}{\sqrt{3/10}} \approx N(0, 1)$.

Vidare är det observerade värdet $z = \dfrac{\frac{107}{30} - 3}{\sqrt{3/10}} = 1.80$ och vi förkastar H_0 om $z > c = z_{0.95} = 1.645$. Alltså, kan $\underline{\underline{H_0 \text{ förkastas på nivån } 5\%}}$.

b) Styrkan för $\lambda = 4$ ges av

$$h(\lambda = 4) = \mathrm{P}\left(\frac{\hat{\lambda} - 3}{\sqrt{3/10}} > 1.645 \text{ då } \lambda = 4\right) = \mathrm{P}\left(\hat{\lambda} > 3.5202 \text{ då } \hat{\lambda} \approx N\left(4, \sqrt{\frac{4}{30}}\right)\right)$$

$$= 1 - \Phi\left(\frac{3.5202 - 4}{\sqrt{4/30}}\right) = 1 - \Phi(-1.314) = \Phi(1.314) = \underline{\underline{0.9056}}$$

vilket är en bra styrka.

6) Vi har att

$$\mathrm{E}(\widehat{P}^2) = \mathrm{var}(\widehat{P}) - \left(\mathrm{E}(\widehat{P})\right)^2 = \frac{p(1-p)}{n} + p^2$$

och

$$\mathrm{E}\left(\frac{\widehat{P}(1-\widehat{P})}{n}\right) = \frac{\mathrm{E}(\widehat{P}) - \mathrm{E}(\widehat{P}^2)}{n} = ... = \underline{\underline{\frac{(n-1)p(1-p)}{n^2}}}$$

vilket är nära $\frac{p(1-p)}{n}$ för stora n.

En väntevärdesriktig skattning av $\mathrm{var}(\widehat{P}) = \frac{p(1-p)}{n}$ är alltså $\underline{\underline{\frac{\widehat{P}(1-\widehat{P})}{n-1}}}$ eftersom $\mathrm{E}\left(\frac{\widehat{P}(1-\widehat{P})}{n-1}\right) = \frac{p(1-p)}{n}$.