Kurskod: TAMS24
Lösningar

Statistisk teori, grundkurs

Lösningsförslag till tentamen måndagen den 18 augusti 2014 kl 8–12.

1) Vi har hypotesen

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = \frac{1}{9}, p_6 = p_7 = \frac{2}{9} \quad \text{mot} \quad H_1 : \text{inte alla } p_i \text{ lika}$$

Vidare gäller att

N_i	6	8	11	7	10	18	30
np_i	10	10	10	10	10	20	20

Teststorhet

$$T = \sum_{i=1}^{7} \frac{(N_i - np_i)^2}{np_i} = \underline{\underline{8.2 < \chi^2_{0.95}(6) = 12.60}}$$

$\underline{\underline{H_0 \text{ kan inte förkastas på nivån } 5\%}}$.

2a) $H_0 : \sigma_1 = \sigma_2$ mot $H_1 : \sigma_1 \neq \sigma_2$

Teststorhet: $v = \frac{s_1^2}{s_2^2} = 2.475$

H_0 förkastas om $v < a$ eller om $v > b$. Den s.v. $V \sim F(11, 11)$ om H_0 är sann och tabell ger $b = \frac{2.98+2.91+2.75+2.69}{4} = 2.83$.

Vidare gäller att $P(V < a) = P(\frac{1}{V} > \frac{1}{a})$. Eftersom $\frac{1}{V} \sim F(11, 11)$ får vi att $\frac{1}{a} = b$ dvs $a = \frac{1}{b} = 0.35$

$\underline{\underline{0.35 < 2.475 < 2.83}}$, så vi kan inte förkasta H_0. $\underline{\underline{\text{Vi har inte kunnat visa att } \sigma_1 \neq \sigma_2}}$.

b) Vi vil visa att $\mu_2 > \mu_1$ dvs att $\mu_1 - \mu_2 > 0$ och gör därför ett nedåt begränsat konfidensintervall för $\mu_1 - \mu_2$.

$\hat{\mu}_1 - \hat{\mu}_2 = \bar{x} - \bar{y} = 0.6425$

Den s.v. $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sqrt{\frac{\sigma^2}{12} + \frac{\sigma^2}{12}})$ där σ^2 skattas med $s^2 = \frac{11s_1^2 + 11s_2^2}{22} = 0.2262$, $s = 0.4756$ och $df = 22$

Hjälpvariabel $\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S/\sqrt{6}} \sim t(22)$ och ger intervallet

$$I_{\mu_1 - \mu_2} = (\bar{x} - \bar{y} - ts/\sqrt{6}, \infty) = \underline{\underline{(0.309, \infty)}}$$

där $t = t_{0.95}(22) = 1.72$.

Alltså, $\mu_1 - \mu_2 > 0.309 > 0$ med stor sannolikhet. $\underline{\underline{\text{Metod B ger med stor sannolikhet sämre genomsnittlig töjbarhet}}}$.

3a) Likelihood-funktionen ges av

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\mu} e^{-x_i/\mu} = \mu^{-n} e^{-\sum_{i=1}^{n} x_i/\mu}$$

$$l(\mu) = \ln L(\mu) = -n \ln \mu - \frac{1}{\mu} \sum_{i=1}^{n} x_i$$

$$l'(\mu) = -\frac{n}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{n} x_i$$

$$l'(\mu) = 0 \quad \text{ger} \quad \mu = \bar{x}$$

$$l''(\mu) = \frac{n}{\mu^2} - \frac{2}{\mu^3} \sum_{i=1}^{n} x_i < 0 \quad \text{för } \mu = \bar{x} \quad \text{dvs max}$$

ML-skattningen är alltså $\underline{\underline{\hat{\mu} = \bar{x}}}$ som har variansen

$\text{var}(\widehat{M}) = \text{var}(\bar{X}) = \frac{1}{n} \text{var}(X) = \underline{\underline{\frac{\mu^2}{n}}}$ eftersom $\text{var}(X) = \mu^2$.

b) Vi har att $\hat{\mu} = \bar{x} = 227.45$

$$p = P(X_i > 500) = \int_{500}^{\infty} \frac{1}{\mu} e^{-x/\mu}\, dx = e^{-500/\mu}$$

som skattas med $\underline{\underline{\hat{p} = e^{-500/\hat{\mu}} = 0.111}}$.

4a) $R^2 = \frac{Q_{REGR}}{Q_{RES}} = \frac{9715.5}{18256.0} = \underline{\underline{53.22\%}}$ visar hur stor andel av variationerna hos Y som förklaras av x och u.

b) Parameter β_2 beskriver den systematiska skillnaden i antal sidor mellan B och A. Bilda konfidensintervall för parametern β_2.

$$I_{\beta_2} = \left(\widehat{\beta}_2 \mp \underbrace{t_{0.975}(21)}_{=2.08} d(\widehat{\beta}_2)\right) = (27.000 \mp 2.08 \cdot 8.233) = \underline{\underline{(9.9, 44.1)}}$$

Bara positiva värden i I_{β_2}. Tidskrift B har i genomsnitt fler sidor än A.

c) Bilda prediktionsintervall för $\boldsymbol{u}' = \begin{pmatrix} 1 & 60 & 1 \end{pmatrix}'$.

Hjälpvariabeln $\frac{Y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{S\sqrt{1 + \boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}}} \sim t(21)$ ger intervallet

$$I_{Y_0} = \left(\underbrace{\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}_{=118.004} \mp \underbrace{t_{0.975}(21)}_{=2.08} s\sqrt{1 + \boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}}\right) = \underline{\underline{(74, 162)}},$$

där variansen σ^2 skattas med $s^2 = \frac{Q_{RES}}{21}$, $s = 20.17$ och $\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u} = 0.103$.

Långt! Modellen beskriver inte data så bra, vilket vi också såg på förklaringsgraden i a).

5a) Teststorhet: $\dfrac{\bar{X}-\mu_0}{\sigma/n} \sim N(0,1)$. $\underline{\underline{\text{Förkasta } H_0 \text{ om } \left|\dfrac{\bar{X}-\mu_0}{\sigma/n}\right| > 2.575.}}$

Styrkefunktionen ges av

$$\begin{aligned}
h(\mu) &= P(\text{förkasta } H_0|\mu \text{ sanna värdet}) \\
&= 1 - P\left(-2.575 < \frac{\bar{X}-\mu_0}{\sigma/n} < 2.575|\mu \text{ sanna värdet}\right) \\
&= 1 - P\left(-2.575 + \frac{\mu_0-\mu}{\sigma/n} < \frac{\bar{X}-\mu}{\sigma/n} < 2.575 + \frac{\mu_0-\mu}{\sigma/n}|\mu \text{ sanna värdet}\right) \\
&= \underline{\underline{1 - \left(\Phi\left(2.575 + \frac{\mu_0-\mu}{\sigma/n}\right) - \Phi\left(-2.575 + \frac{\mu_0-\mu}{\sigma/n}\right)\right)}}
\end{aligned}$$

6) Likelihood-funktionen ges av

$$L(\theta) = \begin{cases} f(x_1)\cdot \ldots \cdot f(x_7) = \left(\frac{1}{2\theta}\right)^7 & \text{om } |x_i - 0.5| \leq \theta \text{ för alla } i, \\ 0 & \text{annars.} \end{cases}$$

$L(\theta)$ har max för $\hat{\theta} = max|x_i - 0.5| = 0.39$.

Alltså, $\underline{\underline{\text{ML-skattningen ges av } \hat{\theta} = 0.39}}$.