

Kurskod: TAMS65

Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningförslag till tentamen torsdagen den 21 mars 2019 kl 14–18

1. Gör ett homogenitetstest och testa

H_0 : Resultaten fördelar sig på samma sätt över de tre nivåerna A_1, A_2, A_3 för samtliga frånvaro kategorier.

mot

H_1 : H_0 ej sann.

Om H_0 är sann har vi $\hat{p}_1 = \widehat{P(A_1)} = \frac{18 + 14 + 3}{99} = \frac{35}{99}$, $\hat{p}_2 = \frac{11 + 12 + 9}{99} = \frac{32}{99}$

och $\hat{p}_3 = \frac{6 + 6 + 20}{99} = \frac{32}{99}$ samt de skattade förväntade frekvenserna $n_i \hat{p}_j$:

| | A_1 | A_2 | A_3 |
|-------|-------|-------|-------|
| A_1 | 12.37 | 11.31 | 11.31 |
| A_2 | 11.31 | 10.34 | 10.34 |
| A_3 | 11.31 | 10.34 | 10.34 |

Teststorhet: $Q = \sum_{i=1}^3 \sum_{j=1}^3 \frac{(N_{ij} - n_i \hat{p}_j)^2}{n_i \hat{p}_j} = \dots = 23.10$.

Den s.v. Q är approx $\chi^2((3-1)(3-1)) = \chi^2(4)$ om H_0 ör sann. H_0 förkastas om $Q > c = \chi_{0.99}^2(4) = 13.28$. Alltså, förkasta H_0 .

Närvaro på undervisning verkar ha betydelse för tentamensresultatet.

2. (a) Teststorhet: $v = \frac{s_3^2}{s_2^2} = 1.77$. H_0 förkastas om $v < a (< 1)$ eller om $v > b (> 1)$.

Den s.v. $V \sim F(20, 20)$ om H_0 är sann. Tabellen ger $b = F_{0.95}(20, 20) = 2.12$. Eftersom $v > 1$ kan inte $v < a$ och eftersom dessutom $1 < v < b$ kan vi inte förkasta H_0 på nivån 10%. Det verkar rimligt att anta $\sigma_1 = \sigma_2 = \sigma_3$.

- (b) Vi konstruerar intervallen $I_{\mu_i - \mu_k}$. Skatta parametrarna $\hat{\mu}_i - \hat{\mu}_k = \bar{x}_i - \bar{x}_k$.

Den s.v. $\bar{X}_i - \bar{X}_k \sim N\left(\mu_i - \mu_k, \sigma \sqrt{\frac{1}{21} + \frac{1}{21}}\right)$ med den sammanvägda

σ^2 -skattningen $s^2 = \frac{20s_1^2 + 20s_2^2 + 20s_3^2}{60} = 14.7564$, $s = 3.8414$, $df = 60$.

Hjälpvariabeln $\frac{\bar{X}_i - \bar{X}_k - (\mu_i - \mu_k)}{S \sqrt{\frac{1}{21} + \frac{1}{21}}} \sim t(60)$ ger intervallen

$$I_{\mu_i - \mu_k} = \left(\bar{x}_i - \bar{x}_k \mp t_{0.995}(60) s \sqrt{\frac{2}{21}} \right) = (\bar{x}_i - \bar{x}_k \mp 3.1534),$$

där $t_{0.995}(60) = 2.66$, dvs.

$$\begin{aligned} I_{\mu_1-\mu_2} &= (-24.4670 \mp 3.1534) = (-27.6 ; -21.3), \\ I_{\mu_1-\mu_2} &= (-29.9337 \mp 3.1534) = (-33.1 ; -26.8), \\ I_{\mu_1-\mu_2} &= (-5.4667 \mp 3.1534) = \underline{\underline{(-8.6 ; -2.3)}}. \end{aligned}$$

Alltså, $2.3 < |\mu_2 - \mu_3| < 8.6$ och åtdragningsmomenten 70 och 80 kan anses likvärda, medan momentet 60 tydligt avviker.

$$3. \quad (a) \quad L(\mu) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\mu} e^{-x_i/\mu} = \frac{1}{\mu^n} e^{-\frac{1}{\mu} \sum_{i=1}^n x_i}$$

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

$$\frac{dl}{d\mu} = -\frac{n}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^n x_i = 0 \text{ ger } \mu = \bar{x}$$

$$\frac{d^2l}{d\mu^2} = \frac{n}{\mu^2} - \frac{2}{\mu^3} \sum_{i=1}^n x_i \Big|_{\mu=\bar{x}} = \frac{n}{\bar{x}^2} - \frac{2}{\bar{x}^3} n\bar{x} = \dots = -\frac{n}{\bar{x}^2} < 0 \Rightarrow \text{d.v.s. max.}$$

Maximum-likelihood-skattningen ges av $\hat{\mu} = \bar{x} = 29.65$.

$$(b) \quad E(\hat{M}) = E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \underbrace{E(X_i)}_{=\mu} = \mu.$$

Ja, \hat{M} är en vvr skattning av μ . Vidare har vi

$$\text{var}(\hat{M}) = \text{var}(\bar{X}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \text{/ober./} = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \underbrace{\text{var}(X_i)}_{=\sigma^2} = \frac{\sigma^2}{n}$$

$$\text{men } X_i \sim \text{Exp}(\mu) \text{ vilket ger att } \text{var}(X_i) = \sigma^2 = \mu^2 \text{ och } \underline{\underline{\text{var}(\hat{M}) = \frac{\mu^2}{n}}}.$$

4. x_1, \dots, x_9 är observationer från $Po(\mu_1)$ och y_1, \dots, y_8 är observationer från $Po(\mu_2)$

Vi söker $I_{\mu_1-\mu_2}$. Skatta parametrarna $\hat{\mu}_1 = \bar{x} = 6.222$ och $\hat{\mu}_2 = \bar{y} = 3.750$.

Vidare gäller att $\sum_{i=1}^9 X_i \sim Po(9\mu_1)$ och $\sum_{i=1}^8 Y_j \sim Po(8\mu_2)$ och eftersom $9\hat{\mu}_1 > 15$ samt $8\hat{\mu}_2 > 15$ är normalapproximation tillåten. Det följer att $\bar{X} - \bar{Y}$ är approx normalfördelad med parametrarna

$$E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2, \quad \text{samt} \quad \text{var}(\bar{X} - \bar{Y}) = \text{var}(\bar{X}) + \text{var}(\bar{Y}) = \frac{\mu_1}{9} + \frac{\mu_2}{8}.$$

$$\text{Hjälpsvariabel} \quad \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\bar{X}}{9} + \frac{\bar{Y}}{8}}} \approx N(0, 1) \quad \text{ger intervallet}$$

$$I_{\mu_1-\mu_2} = \left(\bar{x} - \bar{y} \mp 1.96 \sqrt{\frac{\bar{x}}{9} + \frac{\bar{y}}{8}} \right) = \underline{\underline{(0.36 ; 4.58)}}.$$

Slutsats: Bara positiva värden \Rightarrow felintensiteten verkar vara större för BWR.

5. Den s.v. $X \sim \text{Exp}\left(\frac{1}{\theta^2}\right)$, dvs. $E(X) = \frac{1}{\theta^2}$.

Om H_0 är sann kan vi förvänta oss att $x \approx 1$ medan om H_1 är sann så borde x vara mindre \Rightarrow förkasta H_0 till förmån för H_1 då $x \leq c$, vilket ger

$$0.05 = P(X \leq c | H_0 : \theta = 1) = \int_0^c e^{-x} dx = 1 - e^{-c}$$

$$c = -\ln(0.95) = 0.051.$$

Alltså, det kritiska området är $\{x : x \leq 0.051\}$.

6. (a) Vi vill testa $H_{0i} : \beta_i = 0$ mot $H_{1i} : \beta_i \neq 0$ på nivån 5% för varje $i = 0, 1, 2, 3$.
Teststorhet

$$T_i = \frac{\hat{\beta}_i}{d(\hat{\beta}_i)} = \begin{cases} -0.50, & i = 0, \\ -0.23, & i = 1, \\ 6.24, & i = 2, \\ -0.17, & i = 3. \end{cases}$$

Förkasta H_{0i} om $|T_i| > t_{0.975}(24) = 2.06$, dvs. vi har att $\beta_2 \neq 0$ och de övriga kan vi inte säga något om.

- (b) $R_2^2 = 62.35\% < R_1^2 = 62.49\%$. Förklaringsgraderna är lika eftersom variablerna x_1 och x_3 förklarar väldigt lite av variationen hos datan. Från (a) vet vi ju att x_1 och x_3 inte är signifikanta.

- (c) Betrakta $Y_0 = \beta_0 + 3\beta_2 + \varepsilon_0 = \mathbf{u}'\boldsymbol{\beta} + \varepsilon_0$ där $\mathbf{u}' = (1 \ 3)$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_2 \end{pmatrix}$ och $\varepsilon_0 \sim N(0, \sigma)$. Skatta σ^2 med $s^2 = \frac{Q_{RES}}{26} = \frac{253.09}{26} = 9.73$, dvs. $s = 3.12$.
Skatta $\mathbf{u}'\boldsymbol{\beta}$ med $\mathbf{u}'\hat{\boldsymbol{\beta}} = 2.3819$ som är en observation från den s.v.

$$\mathbf{u}'\hat{\boldsymbol{\beta}} \sim N\left(\mathbf{u}'\boldsymbol{\beta}, \sigma\sqrt{\mathbf{u}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{u}}\right),$$

där $\mathbf{u}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{u} = 0.0481$. Hjälpvariabel $\frac{\mathbf{u}'\hat{\boldsymbol{\beta}} - \mathbf{u}'\boldsymbol{\beta}}{s\sqrt{\mathbf{u}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{u}}} \sim t(26)$ som ger intervallet

$$I_{\beta_0+3\beta_2} = \left(\mathbf{u}'\hat{\boldsymbol{\beta}} \mp \underbrace{t_{0.975}(26)}_{=2.06} s\sqrt{\mathbf{u}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{u}}\right) = \underline{\underline{(0.972 ; 3.791)}}.$$

- (d) $\hat{Y}_4 - \hat{Y}_2 = \hat{\beta}_0 + 4\hat{\beta}_2 - (\hat{\beta}_0 + 2\hat{\beta}_2) = \underline{\underline{2\hat{\beta}_2 = 1.87}}$