Kurskod: TAMS65
Provkod: TEN2

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningsförslag 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 $\widehat{p}_1 = \widehat{P(A_1)} = \dfrac{18+14+3}{99} = \dfrac{35}{99}$, $\widehat{p}_2 = \dfrac{11+12+9}{99} = \dfrac{32}{99}$ och $\widehat{p}_3 = \dfrac{6+6+20}{99} = \dfrac{32}{99}$ samt de skattade förväntade frekvenserna $n_i\widehat{p}_j$:

A_1	A_2	A_3
12.37	11.31	11.31
11.31	10.34	10.34
11.31	10.34	10.34

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

   Den s.v. Q är approx $\chi^2((3-1)(3-1)) = \chi^2(4)$ om H_0 ör sann. H_0 förkastas om $Q > c = \chi^2_{0.99}(4) = 13.28$. Alltså, förkasta H_0.
   Närvaro på undervisning verkar ha betydelse för tentamensresultatet.

2. (a) Teststorhet: $v = \dfrac{s_3^2}{s_2^2} = 1.77$. 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 $\widehat{\mu}_i - \widehat{\mu}_k = \bar{x}_i - \bar{x}_k$.

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

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

   Hjälpvariabeln $\dfrac{\bar{X}_i - \bar{X}_k - (\mu_i - \mu_k)}{S\sqrt{\dfrac{1}{21} + \dfrac{1}{21}}} \sim t(60)$ 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. (a) $L(\mu) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\mu} e^{-x_i/\mu} = \frac{1}{\mu^n} e^{-\frac{1}{\mu}\sum_{i=1}^n x_i}$

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

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

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

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

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

<u>Ja, $\widehat{M}$ är en vvr skattning av μ.</u> Vidare har vi

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

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

(c) Centrala gränsvärdessatsen ger att $\bar{X} \approx N\left(\mu, \frac{\mu}{\sqrt{40}}\right)$ eftersom $\mathrm{var}(X_i) = \sigma^2 = \mu^2$ och hjälpvariabeln $\frac{\bar{X} - \mu}{\mu/\sqrt{40}} \approx N(0,1)$ som ger intervallet

$$I_\mu = \left(\frac{\bar{x}}{1 + \frac{1.96}{\sqrt{40}}}\ ,\ \frac{\bar{x}}{1 - \frac{1.96}{\sqrt{40}}}\right) \approx \underline{\underline{(22.64\ ;\ 42.96)}}.$$

($I_\mu = \left(\widehat{\mu} \mp 1.96\frac{\widehat{\mu}}{\sqrt{40}}\right)$ är ok, även om det inte är lika bra approximation.)

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

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

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 $9\widehat{\mu}_1 > 15$ samt $8\widehat{\mu}_2 > 15$ är normalapproximation tillåten. Det följer att $\bar{X} - \bar{Y}$ är approx normalfördelad med parametrarna

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

Hjälpvariabel $\quad \dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\bar{X}}{9}+\dfrac{\bar{Y}}{8}}} \approx N(0,1) \quad$ 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. $X \sim Exp\left(\dfrac{1}{\theta^2}\right)$, dvs. $\mathrm{E}(X) = \dfrac{1}{\theta^2}$.
   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

$$\begin{aligned} 0.05 = P(X \leq c | H_0 : \theta = 1) &= \int_0^c e^{-x}dx = 1 - e^{-c} \\ c = -\ln(0.95) = 0.051. \end{aligned}$$

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