STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
LÖSNINGAR
22 augusti 2024

Lösningar

Tentamen i Statistisk analys, 22 augusti 2024

Uppgift 1

a) Sant

b) Falskt

c) Sant

d) Falskt

e) Sant

Uppgift 2

a) Under förutsättnng att $\sigma = 0.5$ är känt så ges ett 95% konfidensintervall för μ av

$$\bar{x} \pm \lambda_{0.025} \frac{\sigma}{\sqrt{n}} = 14.26 \pm 1.96 \frac{0.5}{\sqrt{8}} = 14.26 \pm 0.35 = [13.91,\ 14.61].$$

b) Om σ är okänd skattas σ^2 med $s^2 = (n-1)^{-1} \sum_i (x_i - \bar{x})^2 = 0.548^2$, så σ skattas med $s = 0.548$. Konfidensintervallet för μ blir i detta fall

$$\bar{x} \pm t_{0.025}(n-1) \frac{s}{\sqrt{n}} = 14.26 \pm 2.365 \frac{0.548}{\sqrt{8}} = 14.26 \pm 0.46 = [13.81,\ 14.71].$$

Uppgift 3

a) Genom att stoppa in belopen i uttrycken så får vi $\beta^* = S_{xy}/S_{xx} = 1.051$ och $\alpha^* = \bar{y} - \beta^*\bar{x} = -0.294$. 95% konfidensintervall för de två skattningarna ges av (med skattat s används t-fördelning med 4 frihetsgrader i stället)

$\alpha^* \pm \lambda_{0.025}\sigma\sqrt{\sum_i x_i^2/(nS_{xx})} = -0.294 \pm 1.530 = [-1.824, 1.236]$, och

$\beta^* \pm \lambda_{0.025}\sigma/\sqrt{S_{xx}} = 1.051 \pm 0.2963 = [0.755, 1.347]$.

b) Det gäller att $\beta^* \sim N(\beta, \sigma^2/S_{xx})$. För bevis av detta hänvisas till boken s. 432.

Uppgift 4

a) Eftersom några löner sticker ut (framförallt 95 tkr för en australiensiska) så är medelvärde/väntevärde missvisande. Därför bättre att studera rangerna av lönerna (vilket inkluderar medianlönen).

b) Vi använder Wilcoxon's 2-stickprovstest och rangordnar de 12 lönerna ihop. Summan av rangerna för de australienska lönerna blir $R = 1 + 2 + 3 + 5 + 7 + 12 = 30$. Som synes ligger de lägre än svenskornas, men är det signifikant? Om vi väljer signifikansnivå $\alpha = 0.05$ så ska vi förkästa om R ligger signifikant högt eller lågt, med 2.5% på vardera sida. Från tabellen (med $m = n = 6$) får vi därmed att vi ska förkasta om $R \leq 26$ eller om $R \geq 52$. Eftersom inget av detta gäller så förkaster vi inte H_0. Vi kan alltså på detta lilla datamaerial inte utesluta att lönefördelningen är desamma i de två populationerna.

Uppgift 5

Detta är en s.k. kontingenstabell för vilket ett homogenitetstest kan utföras. Om valet av inriktning är detsamma för män och kvinnor så blir det förväntade antalet e_{ij} i respektive cell $n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot}$. T ex blir $e_{11} = 29*23/66 = 10.11$. Med formeln överst i tentan får vi $Q = \sum_{ij}(n_{ij} - e_{ij})^2/e_{ij} = ... = 1.79$. Antalet frihetsgrader blir $(2-1)(3-1) = 2$ eftersom vi har 6 celler men fixar rad och kolumnsummor. Vi ska således förkaste H_0 om $Q > \chi^2_{0.05}(2) = 5.99$ vilket absolut inte gäller. Det finns således ingen anledning att tro att valen för de olika inriktningarna skiljer sig åt mellan könen.

Uppgift 6

a)

Likelihooden blir

$$L(\lambda) = \prod_i \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_i x_i} e^{-\lambda n}}{\prod_i x_i!}.$$

Loglikelihooden blir således

$$\ell(\lambda) = \sum_i x_i/\lambda + n\lambda - \log(\prod_i x_i!).$$

För att finna ML-skattningen så deriverar vi loglikelihooden och får

$$\ell'(\lambda) = \frac{\sum_i x_i}{\lambda} - n, \qquad \ell''(\lambda) = -\frac{\sum_i x_i}{\lambda^2}.$$

Om vi sätter derivatan till 0 får vi således $\hat{\lambda} = \sum_i x_i/n = \bar{x}$. För säkerhets skull kollar vi att andra derivatan är negativ vilket indikerar att vi hittat ett maximum. Det numeriska värdet för våra observationer är $\hat{\lambda} = \bar{x} = (14 + \cdots + 23)/6 = 17$.

b) Medelvärdet av 6 Poissonföredlningar (med hyfsat stora väntevärden) blir approximativt normalfördelat. Variansen för en enskild Poissonfördelning är λ, variansen av medelvärdet blir således $\lambda/6$ och standardavvikelsen för medelvärdet blir således $\sqrt{\lambda/6}$. Ett 95% konfidensintervall ges därför av

$$\hat{\lambda} \pm 1.96 * \sqrt{\hat{\lambda}/6} = 17 \pm 3.30 = [13.7, 20.3].$$