

Lösningar

Tentamen i Statistisk analys, 14 augusti 2025

Uppgift 1

- a) Falskt
- b) Sant
- c) Sant
- d) Sant
- e) Falskt

Uppgift 2

a) Beräkningar ger att $\bar{x} = 3.019$ och $s^2 = (\sum_i x_i^2 - n\bar{x}^2) / 7 = 0.136^2$. Vi får $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = 0.40$. Detta skall jämföras med t-fördelning med 7 frihetsgrader. Inte alls signifikant (oberoende av rimligt signifikansnivå)!

b) Ett 95% konfidensintervall för μ ges av $\bar{x} \pm t_{0.025}(7)s/\sqrt{8} = 3.019 \pm 0.113 = (2.906, 3.132)$.

Uppgift 3

Tvåsidigt Wilcoxons 2-stickprovstest. $m = 5$ och $n = 7$. Om $R =$ rangsumman i mindre stickprovet så ska man förkasta om $R \leq K_-$ eller om $R \geq K_+$ där $K_- = 20$ och $K_+ = 45$ hittas i Tabellen (för 0.025 eftersom testet är tvåsidigt). Vårt $R = 22$ varför hypotesen inte skall förkastas (men nästan). Således är det inte statistiskt säkerställt på 5%-nivån att de två lampsorterna har olika livslängd.

b) Livslängsfördelningen verkar mycket skev, dvs tjock högersvans. En enskild stor observation påverkar därför ett parametriskt test för mycket.

Uppgift 4

a) En lämplig modell är att anta linjär regression, så att $Y = \alpha + \beta x + \epsilon$, där Y är antal IVA-inläggningar, x är vecka och ϵ är slumpfelet. Eftersom Y är heltalsvärd är den inte exakt normalfördelad, ett rimligare antagande är Poisson, men Poisson-fördelningen med stort väntevärde liknar mycket normalfördelningen.

Från de beräknade summorna får man enkelt: $\bar{x} = 49$, $\bar{y} = 167.8$, $S_{xx} = \sum_i x_i^2 - n\bar{x}^2 = 10$, $S_{yy} = \sum_i y_i^2 - n\bar{y}^2 = 456.8$ och $S_{xy} = \sum_i x_i y_i - n\bar{x}\bar{y} = 59$. Från formelsamlingen skattningar får vi $\beta^* = S_{xy}/S_{xx} = 5.9$, $\alpha^* = \bar{y} - \beta^*\bar{x} = -121.3$ och $s^2 = SSE/(n-2) = (S_{yy} - S_{xy}^2/S_{xx})/(n-2) = 6.02^2$.

Ett 95% konfidensintervall för β ges av $\beta^* \pm t_{0.025}(3)s\frac{1}{\sqrt{S_{xx}}} = 5.90 \pm 3.18 * 6.02 * 0.316 = 5.9 \pm 6.05$. Eftersom intervallet innehåller 0 så kan $H_0 : \beta = 0$ inte uteslutas. Ökningen är således inte statistiskt säkerställd (även om det är nära). Jag tycker det är mest rimligt med en två-sidig alternativhypotes varför jag gjorde tvåsidigt konfidensintervall. En som gör ensidig mothypotes gör ensidigt intervall, och då förkastas H_0 på denna signifikansnivå.

b) Ett 95% prediktionsintervall för antal inläggningar vecka 52 ges av $I_{Y_{52}} : \alpha^* + \beta^* * 52 \pm t_{0.025}(3)s\sqrt{1 + 1/n + (52 - \bar{x})^2/S_{xx}} = 185.5 \pm 3.18 * 6.02 * 1.45 = 185.5 \pm 27.7 = (157, 214)$ (bäst att välja närmaste heltal utanför sifferintervallet).

Uppgift 5

a) Data kan skrivas som en kontingenstabell med O_{i1} = observerat antal kunder utan skada i stock i och $O_{i2} = 100 - O_{i1}$ observerat antal kunder med skada i stock i . Under H_0 om att alla kundstockar är lika blir förväntat antal i respektive cell $e_{i1} = n_i \frac{O_{.1}}{n_{..}} = 100 \frac{162}{400} = 40.5$ och $e_{i2} = n_i \frac{O_{.2}}{n_{..}} = 100 \frac{238}{400} = 59.5$.

Nollhypotesen testas genom att jämföra $X \sum_{ij} (O_{ij} - e_{ij})^2 / e_{ij}$ med χ^2 -fördelningen med $(4-1)(2-1)=3$ frihetsgrader.

Vi får

$$X = \frac{(32 - 40.5)^2 + (47 - 40.5)^2 + (39 - 40.5)^2 + (44 - 40.5)^2}{40.5} + \frac{(68 - 59.5)^2 + (53 - 59.5)^2 + (61 - 59.5)^2 + (56 - 59.5)^2}{59.5} = 5.35.$$

Vi har att $\chi_{0.95}^2(3) = 7.81$ så eftersom $5.35 < 7.81$ så kan vi inte förkasta H_0 om att alla kundstockar är jämbördiga.

b) Om man tittar i χ^2 -tabellan för 3 frihetsgrader ser man att $\chi_{0.75}^2(3) = 4.11$ och $\chi_{0.9}^2(3) = 6.25$ och 5.31 ligger emellan dessa värden, lite närmre det senare. En rimlig skattning är därför att det ungefär svarar mot $\chi_{0.85}^2(3)$ vilket betyder att p -värdet är ungefär 0.15.

Uppgift 6

a) Likelihooden ges av $L(\theta) = \theta^n \prod_i x_i^{\theta-1}$. Loglikelihooden ges av $\ell(\theta) = n \log(\theta) + (\theta - 1) \sum_i \log(x_i)$. Om man deriverar med avseende på θ får man $\ell'(\theta) = n/\theta + \sum_i \log(x_i)$. Om man sätter detta till 0 får vi $\hat{\theta} = -n / \sum_i \log(x_i)$.

b) För angivna data blir ML-skattingen $\hat{\theta} = 2.42$.