

Lösningar

Tentamen i Statistisk analys, 14 januari 2026

Uppgift 1

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

Uppgift 2

a) Eftersom $\bar{X} - \bar{Y}$ under nollhypotesen om att de två växtextrakten har samma väntevärde är normalfördelat med väntevärde 0 och varians $\sigma^2(1/n_1 + 1/n_2)$ blir referensvariabeln $z = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{1/n_1 + 1/n_2}} = \frac{3.7 - 4.3}{1 \cdot \sqrt{1/6 + 1/6}} = -1.04$. Under nollhypotesen är denna referensvariabel standard normal. Vi ska således inte förkasta nollhypotesen (eftersom $|z| = 1.04 < 1.96$).

b) Skattningarna av varianserna blir $s_x^2 = 0.948$ respektive $s_y^2 = 0.704$, så den polade variansen blir $s_p^2 = (s_x^2 + s_y^2)/2 = 0.826$ eftersom de två stickproven är lika stora. Standardavvikelsen blir således $s_p = 0.909$. Referensvariabeln blir därför $t = \frac{\bar{x} - \bar{y}}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{3.7 - 4.3}{0.909\sqrt{1/6 + 1/6}} = -1.14$. Detta skall jämföras med t -fördelningen med $n_1 + n_2 - 2 = 10$ frihetsgrader. Slutsatsen blir att man inte kan förkasta H_0 om att de två växtextrakten är ekvivalenta (gränsen går vid $t_{0.025}(10) = 2.23$).

Uppgift 3

a) Vid beräkning får man: $n = 5$, $\bar{x} = 1.5$, $\sum_i x_i^2 = 13.75$, $S_{xx} = 2.5$, respektive $\bar{y} = 36.6$, $\sum_i y_i^2 = 7763$, $S_{yy} = 1065.2$, samt $\sum_i x_i y_i = 326$ vilket ger $S_{xy} = 51.5$. Temperaturökningen per timme är exakt lutningskoefficienten β . Detta skattas med $\beta^* = S_{xy}/S_{xx} = 20.6$. Felkvadratssumman ges av $SSE = S_{yy} - S_{xy}^2/S_{xx} = 4.3$, så standardavvikelsen skattas med $s = \sqrt{SSE/(n-2)} = 1.20$. Ett 95% konfidensintervall för β ges av $\beta^* \pm t_{\alpha/2}(n-2) \cdot s/\sqrt{S_{xx}} = 20.6 \pm 3.182 \cdot 1.20/\sqrt{2.5} = 20.6 \pm 2.4$.

Medeltemperaturen i skinkan när den sattes in är ju medeltemperaturen vid $x = 0$, dvs interceptet $\alpha + \beta \cdot 0 = \alpha$. Detta skattas med $\alpha^* = \bar{y} - \beta^* \bar{x} = 36.6 - 20.6 \cdot 1.5 = 5.7$. Konfidensintervallet ges av $\alpha^* \pm t_{\alpha/2}(n-2) \cdot s \sqrt{\frac{\sum_i x_i^2}{n S_{xx}}} = 5.7 \pm 4.0$.

b) Om vi vill prediktera vad temperaturen blir vid mätning vid $x_0 = 3$ ges punktskattningen av $\alpha^* + \beta^* x_0 = 67.5$ och prediktionsintervallet blir $\alpha^* + \beta^* x_0 \pm t_{\alpha/2}(n-2) \cdot s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = 67.5 \pm 5.5$.

Uppgift 4

Vi har en kontingenstabell som vi vill göra homogenitetstest för, dvs testa om de olika ländernas doktorander är lika bra och därmed har samma fördelning över de tre betygen. Under denna hypotes ges förväntat antal i respektive cell av $e_{ij} = n_{i.} n_{.j} / n$. Motsvarande värden blir

Land Betyg	U	G	VG	Totalt
Sverige	9.5	15	9.5	34
Norge	9.5	15	9.5	34
Totalt	19	30	19	68

Vi beräknar $\chi_{obs}^2 = \sum_{i,j} (o_{ij} - e_{ij})^2 / e_{ij} = 1.92$. Detta ska jämföras med en χ^2 -fördelning med $(2-1)(3-1) = 2$ frihetsgrader. Denna fördelning har väntevärde 2, så det observerade värdet är fullt rimligt och det finns ingen anledning alls att påstå att Sverige är signifikant bättre än Norge baserat på detta resultat (gränsen att förkasta går vid $\chi_{0.05}^2(2) = 5.99$).

Uppgift 5

a) Likelihooden ges av

$$L(\beta) = \prod_i \beta e^{-\beta x_i} = \beta^n e^{-\beta \sum_i x_i}, \quad (\text{om alla } x_i > 0).$$

Loglikelihooden blir således $\ell(\beta) = n \log \beta - \beta \sum_i x_i$. Vi söker maximum med avseende på β . Vi deriverar och får $\ell'(\beta) = n/\beta - \sum_i x_i$ och andra derivatan blir $\ell''(\beta) = -n/\beta^2 < 0$ så

när derivatan är 0 erhålls maximum. Lösningen till $\ell'(\beta) = 0$ ges av $\hat{\beta} = n/\sum_i x_i = 1/\bar{x} = 1/3.2 = 0.31$.

b) För exponentialfördelningen gäller $E(X) = 1/\beta$. Det värde på $1/\beta$ som maximerar likelihooden är såklart 1 dividerat med det värde som maximerar likelihooden med avseende på β . Således blir ML-skattningen $\widehat{E(X)} = 1/\hat{\beta} = \bar{x} = 3.2$, dvs medelvärdet av observationerna.

c) Skattningen i b) var $\bar{x} = \sum_{i=1}^{10} x_i/10$. Variansen av denna skattning blir $V(\bar{X}) = V(\sum_i X_i)/100 = 10V(X)/100 = V(X)/10$. Variansen för en enskild exponentialfördelad variabel är $1/\beta^2$, så standardavvikelsen för $\bar{X}_{10}$ blir $1/(\beta\sqrt{10})$. Tack vare centrala gränsvärdessatsen så är $\bar{X}_{10}$ approximativt normalfördelad med väntevärde $E(X) = 1/\beta$ och standardavvikelse $1/(\beta\sqrt{10})$. Ett 95% konfidensintervall för $E(X) = 1/\beta$ blir således $1/\hat{\beta} \pm 1.96 * 1/(\hat{\beta}\sqrt{10}) = 3.2 \pm 1.98 = [1.2, 5.2]$. Ett ganska brett konfidensintervall vilket beror på att exponentialfördelningen innehåller ganska mycket slump.

Uppgift 6

Om vi rangordnar åldrar och löner så blir dessa enligt följande (u, v) : (1,2), (2,1), (3,4), (4,3), (5,6), (6,5), (7,8), (8,7), (9,10), (10,9). Om vi tittar på $D_i = (u_i - v_i)$ så blir den ± 1 för alla i , så $\sum_i D_i^2 = 10$. Vi har $n = 10$ observationer, så Spearmankorrelationen blir $r_S = 1 - 6 \sum_i D_i^2 / (n(n^2 - 1)) = 0.94$.

Under H_0 så är r_S approximativt normalfördelad med väntevärde 0 och standardavvikelse $1/\sqrt{n-1} = 1/3$. Vår teststatistika blir således $z = r_S/(1/3) = 2.82$. Detta är statistiskt signifikant vare sig vi betraktar ett tvåsidigt test (gräns 2.58) eller ensidigt (gräns 2.33). Det råder alltså ingen tvekan om att vd:ars löner i Svenska storbolag har ett samband med deras ålder.