STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
LÖSNINGAR
9 januari 2023

Lösningar

Tentamen i Statistisk analys, 9 januari 2023

Uppgift 1

a) Falskt

b) Sant

c) Falskt

d) Falskt

e) Sant

Uppgift 2

a) $\bar{x} = 3.7, \bar{y} = 4.3, \sum_i x_i^2 = 86.88, \sum_i y_i^2 = 114.46.$

b) Eftersom $\bar{X} - \bar{Y}$ under nollhypotesen (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 $|z| = 1.04 < 1.96$).

c) 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 således $t = \frac{\bar{x}-\bar{y}}{s_p\sqrt{1/n_1+1/n_2}} = \frac{3.7-4.3}{0.909\sqrt{1/6+1/6}} = -1.14$. Detta skall jämföras med

t-fördelningen med $n_1 + n_2 - 2 = 10$ frihetsgrader. Slutsatsen blir således att man inte kan förkasta H_0 om att de två växtextrakten är ekvivalenta.

Uppgift 3

a) Temperaturökningen per timme är detsamma som lutningskoefficienten β. Denna skattas med $\beta^* = S_{xy}/S_{xx} = 20.6$. Felkvadratsumman 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 $\alpha^* \pm t_{\alpha/2}(n-2) \cdot s\sqrt{\frac{\sum_i x_i^2}{nS_{xx}}} = 5.7 \pm 4.0$.

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 $\alpha^* + \beta^* t_0 \pm t_{\alpha/2}(n-2) \cdot s\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{S_{xx}}} = 67.5 \pm 5.5$.

Uppgift 4

a) Data verkar inte komma från en normalfördeling utan en klart skev fördelning. Detta syns genom att två av observationerna är väsentligt större än övriga. Dessa två observationer skulle ha mycket stort genomslag om medelvärdena jämfördes. Således bättre med icke-parametrisk metod.

b) Lämplig icke-parametrisk metod är Wilcoxons 2-stickprovstest. Sorterade i storleksordning blir x-värdena respektive y-värdena

1.3, 1.9, 2.4, 3.5, 6.1, 13.6

2.9, 4.0, 4.1, 5.4, 6.2, 6.5, 7.4, 8.3, 16.8

Om dessa rangordnas gemensamt blir således rangerna för x-värdena respektive y-värdena:

1, 2, 3, 5, 9, 14

4, 6, 7, 8, 10, 11, 12, 13, 15.

Rangsumman för det mindre stickprovet (med $n_1 = 6$ obs) blir således $R = 34$. På 95% säger tabellen (med $n_1 = m = 6$ och $n_2 = n = 9$ att vi ska förkasta H_0 : de två stickproven har samma fördelning, om $R \leq 31$ eller $R \geq 65$. Då inget av detta gäller blir slutatsen att vi inte kan förkasta H_0.

Uppgift 5

a) Det gäller att $E(X) = 1/\beta$. Momentmetoden säger att vi ska skatta β genom att lösa ekvationen $E(X) = \bar{x}$, dvs $1/\beta = \bar{x}$. Lösningen blir således $\beta^* = 1/\bar{x} = 1/3.54 = 0.282$.

b) Likelihooden ges av $L(\beta) = \prod_i f(x_i;\beta) = \prod_i \beta e^{-\beta x_i} = \beta^n e^{-beta \sum_i x_i}$. Loglikelihooden blir således $\ell(\beta) = n \log \beta - \beta \sum_i x_i$. För att finna maximum dervierar vi och sätter derivatan till noll. Detta ger $\ell'(\beta) = n/\beta - \sum_i x_i$, så $\beta = 1/\bar{x} = 0.282$, dvs samma skattningen som momentmetoden. För säkerhets skulle deriverar vi en gång till för att försäkra sig om att det var ett maximum och inte minimum: $\ell''(\beta) = -n/\beta^2 < 0$.

Uppgift 6

a) Vi genomför ett homogenitetstest. Förväntade värden i tabellen under antagande att frekvensen på olika arter är densamma på de tre öarna, blir $e_{ij} = n\hat{p}_i * \hat{q}_j$, där p_i = total andel obs på ö i som alltså är $1/3$ för alla tre öar, och q_j = total andel observationer av art i.

Vi beräknar $Q = \sum_{i,j} (o_{ij} - e_{ij})^2 / e_{ij} = 9.44$. Under H_0 så är $Q \sim \chi^2((r-1)(c-1)) = \chi^2(6)$. Vi förkastar H_0 om $q_{obs} > \chi^2_{0.05}(6) = 12.6$. Eftersom så ej är fallet så kan vi inte förkasta H_0. Det föreligger ingen signifikant skillnad i artfrekvenser på de tre öarna.

b) Givet att vi observerat 120 fåglar så är X = antal starar $\sim Bin(n = 120, p)$. Ett 99% konfidensintervall för p ges således av

$$\hat{p} \pm \lambda_{0.005} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.275 \pm 2.576 * \sqrt{\frac{0.275 * 0.725}{120}} = [0.17,\ 0.38].$$