Lösningar

Tentamen i Statistisk analys, 22 augusti 2023

Uppgift 1

a) Falskt

b) Sant

c) Falskt

d) Sant

e) Sant

Uppgift 2

a) Vi antar på goda grunder att vikterna kan approximeras ha oberoende normalfördelade vikter. Vi får $\bar{x} = 20.17$ och $s = 0.271$. Ett 95% konfidensintervall ges således av $\bar{x} \pm t_{0.05}(9)s/\sqrt{10} = 20.17 \pm 1.8331 * 0.271/\sqrt{10} = (20.01,\ 20.33)$.

b) Vår teststatistika blir

$$t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{10}} = \frac{20.17 - 20}{0.271/\sqrt{10}} = 1.984.$$

Vi förkastar $H_0 : \mu = 20$ till förmån för $H_A : \mu > 20$ på 95%-nivån om $t_{obs} > t_{0.05}(9) = 1.8331$. Eftersom detta gäller $(1.984 > 1.8331)$ så förkastar vi H_0. Slutsatsen är att säckernas genomsnittliga vikt signifikant överstiger 20.00 kg.

Uppgift 3

a) Vi änvänder linjär regression. Hur mycket medeltempersaturen förändras per lattitud är exakt detsamma som värdet på β eftersom $E(Y(x) - Y(x - 1) = \alpha + \beta x - (\alpha + \beta(x-1)) = \beta$. Paramerern β skattas med

$$\beta^* = \frac{S_{xy}}{S_{xx}} = \frac{\sum_i x_i y_i - n^{-1}\sum_i x_i \sum_i y_i}{\sum_i x_i^2 - n^{-1}(\sum_i x_i)^2} = -0.723$$

Stickprovsvariansen skattas med

$$s^2 = \frac{SSE}{n-2} = \frac{1}{9}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = 9^{-1}\left(68.3054 - \frac{81.59^2}{112.78}\right) = 1.0312,$$

så $s = 1.016$.

Antalet frihetsgrader är $n - 2 = 9$ och $\alpha = 0.01$ så $t_{\alpha/2}(n-2) = 3.25$. Ett 99% konfidensintervall för β ges således av $\beta^* \pm t_{\alpha/2}(n-2)\frac{s}{\sqrt{S_{xx}}} = -0.723 \pm 0.311 = [-1.034,\ -0.412]$.

b) Det gäller att $R^2 = 1 - \frac{SSE}{SST} = \frac{S_{xy}^2}{S_{xx}S_{yy}} = 0.864$.

c) En skattning av den förväntade medeltemperaturen i Bollnäs (med lattitud $x_0 = 61.34$) ges av $\alpha^* + \beta^* x_0 = \bar{y} - \beta^*\bar{x} + \beta^* x_0 = \bar{y} + \beta^*(x_0 - \bar{x}) = 5.936 - 0.723(61.34 - 59.73) = 4.79$. Den förväntade medeltemperaturen i Bollnäs är således 4.79 grader.

Uppgift 4

a) Likelihooden ges av $L(\lambda) = \prod_i \lambda^{x_i} e^{-\lambda}/x_i!$ vilket ger log-likelihooden $\ell(\lambda) = \log\lambda \sum_i x_i - n\lambda - \sum_i \log(x_i!)$. Derivering ger $\ell'(\lambda) = \sum_i x_i/\lambda - n$. Derivetan sätts till noll ger ML-skattningen $\hat{\lambda} = \sum_i x_i/n = \bar{x}$. Om man dervierar en gång till får man $\ell''(\lambda) = -\sum_i x_i/\lambda^2 < 0$ vilket visar att det är et tmaximum vi erhållit. De numeriska skattningarna blir således $\hat{\lambda}_1 = \bar{x} = 4$ och $\hat{\lambda}_2 = \bar{y} = 4$.

b) Vi "glömmer" Poissonfördelningen och testar $H_0 : \mu_i - \mu_2 = 0$ (där väntevärdet nu skrivs som μ för att mer likna vanliga teorin). Den polade variansen blir $s_p^2 = 1.44$ (det går även att skatta lite mer krångligt där de två stickproven tillåts ha olika varianser). Under nollhypotesen gäller att $\bar{X} - \bar{Y}$ är normalfördelad med väntavärde 0 och varians $(1/n_1 + 1/n_2)\sigma^2$, så teststatistikan blir $t_{obs} = (\bar{x} - \bar{y})/s_p\sqrt{1/10 + 1/10} = 1.86$. Om vi t ex testar på 95%-nivån blir slutsatsen att vi inte kan förkasta H_0 eftersom $|t_{obs} <$

$t_{0.025}(18) = 2.107$. Vi kan alltså inte säkert hävda att de två arterna har olika genomsnittligt antal ungar.

Uppgift 5

a) Eftersom de möjliga utfallen är ganska få, men framför allt för att fördelningen verkar vara skev med några få väldigt höga värden så kan normalfördelningen absolut ifrågasättas. Det kan t.o.m. vara så att antal partners har ett väntevärde som är nästintill oändligt. I sådana fall är test av väntevärden olämpliga.

b) Alla observationer med samma värde ges samma rang, och den rangen bestäms så att summan av rangerna görs oförändrad. T ex är de fyra minsta observationerna i datamaterialet alla 0. Eftersom dessa ska ges rang 1 till 4 ges alla samma rang $(1 + 2 + 3 + 4)/4 = 2.5$.

c) Vi använder oss av Wilcoxons 2-stickprovs test (även kallat Mann-Whitney). Alla observationer rangordnas och summan av rangerna för det mindre stickprovet adderas. Detta ger för männen $r_{obs} = 76.5$. Antal obs för männen är $m = 8$ och för kvinnor $n = 10$. Man förkastar hypotesen om identiska fördelningar om $R \leq k^-$ eller om $R \geq k^+$, där k^- och k^+ är 2.5% resp 97.5% kvantiler som man finner i tabell. Från tabellen ser man att $k^- = 53$ och $k^+ = 99$ och eftersom 76.5 ligger emellan dessa värden så förkastar vi inte H_0. I själva verket ligger r_{obs} väldigt nära det förväntade värden $E(R) = 76$.

Uppgift 6

a) Y är $Bin(n = 10, p)$ och likelihooden ges av $L(p) = \binom{10}{7}p^7(1-p)^3$. Loglikelihooden blir $\ell(p) = const + 7\log(p) + 3\log(1-p)$. Derivera, sätt derivatan till 0 och lös för p ger ML-skattningen $\hat{p} = 7/10 = 0.7$.

b) Vi vet att apriorifördelningen är proprtionell mot likelhooden multiplicerat med apriorifördelningen:

$f(p|y = 7) \propto f(p) * L(p) = 1 * \binom{10}{7}p^7(1-p)^3 \propto p^7(1-p)^3$ för $p \in [0, 1]$ och annars är $f(p|y = 7) = 0$.

c) Aposteriorifördelningen ovan passar in på beskrivningen av beta-fördelningen ovan för $\alpha = 8$ och $\beta = 4$. Således är aposteriofördelningen $Beta(8, 4)$. Om vi hade haft någon annan beta-fördelad apriorifördelningen så kommer den in i aposteriofördelningen som $constp^{\alpha-1}(1-p)^{\beta-1}$, som när det multipliceras likelihooden ger $constp^{\alpha+7-1}(1-p)^{\beta+3-1}$. Aposteriofördelningen blir således $Beta(\alpha+7, \beta+3)$. Så med Beta-fördelad apriofördelning så blir aposteriofördelningen också beta-fördelad. Beta-fördelningen är därme *konjugatfördelning* till binomialfördelningen (som likelihooden har).