STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MS 4001
LÖSNINGAR
11 februari 2020

Lösningar

Tentamen i Statistisk analys, 11 februari 2020

Uppgift 1

a) Sant

b) Falskt

c) Sant

d) Falskt

e) Falskt

Uppgift 2

a) Vi antar på goda grunder att den enskilda tempereturerna kan antas vara oberoende och normalfördelade. Vi får $\bar{x} = 20.17$ och $s = 0.271$. Ett 90% 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 medeltemperaturen signifikant överstiger 20.00 grader.

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) Det linjära sambandet skattas med Pearsons korrelationskoefficient $\rho^* = r_{xy} = S_{xy}/\sqrt{S_{xx}S_{yy}} = -0.935$. Teststatistikan ges av

$$\frac{\sqrt{n-2} * r_{xy}}{\sqrt{1 - r_{xy}^2}} = -7.45.$$

Detta ska jämföras med t-fördelningen med $n-2 = 8$ frihetsgrader vilket ger väldigt stark signifikans. Tabellen går bara ner till 0.0005 med t-värde 5.04, så klart mindre p-värde än 0.001 (vid tvåsidig mothypotes).

b) Ordningssambandet mäts med Spearmans rangkorrelation. Genom att rangordna y-variablerna beräknar man r_S som ges i formelsamling och blir $r_s = -0.927$ och motsvarande teststatistika blir $t_{obs} = \sqrt{n-1}r_s = -2.78$. Även detta är signifikant då $p_{obs} = P(|Z| > 2.78) = 0.0054$ (vid tvåsidig mothypotes). I båda fallen finns ett starkt negativt samband. Det linjära sambandet är dock ännu starkare än ordningssambandet.

Uppgift 5

a) Likelihooden blir $\prod_i \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$ och loglikelihooden således $\ell(\lambda) = -n\lambda + \log(\lambda)\sum_i x_i - \sum_i \log(x_i!)$. Denna är kontinuerlig och deriverbar, så ML-skattningen ges av lösningen till $\ell'(\lambda) = 0$ (om det bara finns en lösning och andraderivatan är negativ i punkten). Man får

$$\ell'(\lambda) = -n + \frac{\sum_i x_i}{\lambda}.$$

Om denna sätts till 0 får vi $\hat{\lambda} = \bar{x}$. För data ges medelvärdet av $\bar{x} = 9.2$, så detta blir ML-skattningen för genomsnittligt antal dödliga trafikolycker i Stockholms län ett enskilt år.

b) ML-skattningen kan skrivas som $\hat{\lambda} = \sum_{i=1}^{10} X_i/10$, dvs ett medelvärde av ett antal Poissonvariabler. Enligt centrala gränsvärdessatsen blir detta approximativt normalfördelat. Väntevärdet ges av väntevärdet av de enskilda observationerna, dvs λ, och eftersom varje observation har varians λ (en egenskap hos Poissonfördelningen) blir variansen för $\hat{\lambda}$ λ/n där $n = 10$. Standardavvikelsen blir således $\sqrt{\lambda/n}$ vilket kan approximeras med $\sqrt{\hat{\lambda}/n} = 0.956$. Ett 95% konfidensintervall för λ ges således av

$$\hat{\lambda} \pm z_{\alpha/2}\sqrt{\hat{\lambda}/n} = 9.2 \pm 1.96 * 0.938 = 9.2 \pm 1.88 = [7.32, 11.08].$$

Uppgift 6

a) Aposteriofördelningen för μ, $f(\mu|x_1,\ldots,x_n)$ uppfyller $f(\mu|x_1,\ldots,x_n) \propto f(\mu)L(\mu)$, där $f(\mu)$ är apriorifördelningen och $L(\mu)$ är likelihooden. Vi är således intresserad av hur denna produkt beror på μ. Om vi utelämnar faktorer som inte beror av μ får vi

$$f(\mu)L(\mu) \propto e^{-\frac{1}{2}\left(\frac{\mu-3}{1}\right)^2} * e^{-\frac{1}{2}\sum_i\left(\frac{x_i-\mu}{\sigma}\right)^2}.$$

Om vi studerar exponenten, utelämnar faktorn $-1/2$ och bara inkluderar termer som innehåller μ så blir detta:

$$\begin{aligned}\mu^2 - 6\mu + \frac{2\mu\sum_i x_i}{\sigma^2} + \frac{n\mu}{\sigma^2} &= \left(1+\frac{n}{\sigma^2}\right)\mu^2 - \left(6 + \frac{2n\bar{x}}{\sigma^2}\right)\mu \\ &= \left(1+\frac{n}{\sigma^2}\right)\left(\mu^2 - 2\left(\frac{3+\frac{n}{\sigma^2}\bar{x}}{1+\frac{n}{\sigma^2}}\right)\mu\right).\end{aligned}$$

Detta är på formen $a\left(\mu^2 - 2b\mu\right)$, så med lämplig konstant-term (dvs oberoende av μ) adderad inne i parentesen, närmare bestämt b^2, ser vi att uttrycket kan skrivas som $\left(\frac{(\mu-b)^2}{1/a}\right)$. Från detta drar vi slutsaten att μ måste vara normalfördelad med väntevärde b and varians a, dvs med väntevärde $(3 + n\bar{x})/\sigma^2/(1 + n/\sigma^2)$ och varians $1/(1 + n/\sigma^2)$. Motsvarande siffervärden blir 3.48 respektive 0.034 (så standardavvikelsen blir 0.19). Aposteriorifördelningen är således normalfördelad med väntevärde 3.48 och standardavvikelse 0.19. Således väger det observerade medelvärdet 3.5 klart tyngre än apriorifördelningens väntevärde på 3.0 i aposteriorifördelningen

b) En rimlig punktskattning ges av apoeriorifördelningens väntevärde, dvs $\hat{\mu} = 3.48$.