STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
LÖSNINGAR
12 januari 2022

Lösningar

Tentamen i Statistisk analys, 12 januari 2022

Uppgift 1

a) Sant

b) Sant

c) Sant

d) Sant

e) Falskt

Uppgift 2

a) Under förutsättnng att $\sigma = 0.5$ är känt så ges ett 95% konfidensintervall för μ av

$$\bar{x} \pm \lambda_{0.025}\frac{\sigma}{\sqrt{n}} = 14.26 \pm 1.96\frac{0.5}{\sqrt{8}} = 14.26 \pm 0.35 = [13.91,\ 14.61].$$

b) Om σ är okänd skattas σ^2 med $s^2 = (n-1)^{-1}\sum_i (x_i - \bar{x})^2 = 0.548^2$, så σ skattas med $s = 0.548$. Konfidensintervallet för μ blir i detta fall

$$\bar{x} \pm t_{0.025}(n-1)\frac{s}{\sqrt{n}} = 14.26 \pm 2.365\frac{0.548}{\sqrt{8}} = 14.26 \pm 0.46 = [13.81,\ 14.71].$$

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) Vi antar på goda grunder att vikterna kan approximeras ha oberoende normalfördelade vikter. Vi får $\bar{x} = 20.17$ och $s = 0.271$.

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 tryffelpakaten har vikt som signifikant överstiger 20.0 g.

b) Ett 90% konfidensintervall ges av $\bar{x} \pm t_{0.05}(9)s/\sqrt{10} = 20.17 \pm 1.8331 * 0.271/\sqrt{10} = (20.01,\ 20.33)$.

Uppgift 5

a) Vi väljer 2-sidig hypotes och $\alpha = 5\%$ felrisk, och testar således om socialdemokraterna ligger oförändrat ($p = 0,282$) mot alternativet att en ändring har skett. Vi bildar ett 95% konfidensintervall för p genom att först skatta p med $\hat{p} = X/n = 0,292$, där X = antal som sagt sig vilja rösta på socialdemokraterna och $n = 1500$ antal svarande i undersökningen. X är således $Bin(n = 1500, p)$ som har varians $np(1-p)$. $\hat{p} = X/n$ har således varians $p(1-p)/n$ och standardavvikelse $\sqrt{p(1-p)/n}$, vilket skattas genom att ersätta p med dess skattning $\hat{p} = 0,292$ instoppat för p. Ett 95% konfidensintervall ges därför av

$$\hat{p} \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,292 \pm 1,96 * 0,0117 = [0,269,\ 0,315].$$

Eftersom 28,2% ingår i intervallet så kan inte nollhypotesen förkastas. Dvs den observerade ökning sedan valet på 1% är inte statistiskt säkerställd utan kan mycket väl förklaras av slumpen.

b) Skillnaden mellan de två tidpunkterna för opinionsmätningarna $p_{dec} - p_{okt}$ skattas med $\hat{p}_{dec} - \hat{p}_{okt} = 0,292 - 0,257 = 0,035$, alltså en uppgång på 3.5 procentenheter. Denna skattning har varians $p_{dec}(1-p_{dec})/n + p_{okt}(1-p_{okt})/n$. Ett 95% konfidensintervall ges således av

$$\hat{p}_{dec} - \hat{p}_{okt} \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}_{dec}(1-\hat{p}_{dec})}{n} + \frac{\hat{p}_{okt}(1-\hat{p}_{okt})}{n}} = 0,035 \pm 0,032.$$

Eftersom 0, som svarar mot ingen förändring, *inte* ingår i konfidensintervallet så har socialdemokraterna gjort en statistiskt säkerställd framgång jämfört med i oktober.

Uppgift 6

a) Medianen $x_{0.5}$ skattas med medianen på stickprovet som blir $(x_{(5)} + x_{(6)})/2 = (9,0 + 10,8)/2 = 9,9$.

b) Lämpliga konfidensintervall är $[x_{(1)},\ x_{(10)}]$, $[x_{(2)},\ x_{(9)}]$, osv. Vi väljer det smalaste intervallet som fortfarande har minst 95% konfidensgräns. Vi testar med $[x_{(2)},\ x_{(9)}]$. Det gäller att $P(X_{(2)} < x_{0.5} < X_{(9)}) = 1 - P(X_{(2)} > x_{0.5}) - P(X_{(9)} < x_{0.5})$. Av symmetriskäl är de två negative termerna lika. Den första negativa termen betyder att den näst minsta observationen ska vara större än medianen. Om vi låter Y ange antal observationer som är större än medianen så är detta är ekvivalent med att minst 9 observationer är större än medianen, dvs att $Y \geq 9$. Varje observation är större än medianen med slh 0,5, så $y \sim Bin(n = 10, p = 0,5)$, och vi får $P(Y \geq 9) = \binom{10}{10} 0,5^{10} + \binom{10}{9} 0,5^{10} = 0,0107$. Konfidensgraden blir således $1 - 2 * 0,0107 = 0,979$.

Om vi provar snäppet smalare intervall, dvs $[x_{(3)},\ x_{(8)}]$ kommer detta ha klart för liten konfidensgrad. En av de negativa termerna blir i detta fall $P(Y \geq 8) = \left(\binom{10}{10} + \binom{10}{9} + \binom{10}{8}\right) 0,5^{10} = 0,055$, så konfidensgränsen blir $1 - 2 * 0.055 = 0,89$.

Det 95% konfidensintervallet för medianen $x_{0,5}$ ges således av $[x_{(2)},\ x_{(9)}] = [3,9,\ 23,8]$, och detta intervall har konfidensgräns 95,7%.