STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
LÖSNINGAR
16 februari 2022

Lösningar

Tentamen i Statistisk analys, 16 februari 2022

Uppgift 1

a) Falskt

b) Falskt

c) Sant

d) Sant

e) Sant

Uppgift 2

a) Vi beräknar $t = (\bar{x} - 10)/(s/\sqrt{n}) = 1.857$. Vår referensvariabel har under $H_0 : \mu = 10$ en t-fördelning med $n - 1 = 9$ frihetsgrader. Vi förkastar H_0 om $|t|$ är stor. Från Tabell ser vi att vi ska förkasta H_0 på 95%-nivån om $|t| > 2.26$. Slutsatsen blir således att vi inte kan förkasta H_0. Visserligen är medelvärdet $\bar{x} = 10.32$ en bit ifrån hypotesvärdet 10 men med tanke på datamaterialets spridning kan denna avvikelse gott och väl ske av ren slump.

b) Ett 99% konfidensintervall för μ ges av: $\bar{x} \pm t_{\alpha/2}(n-1)s/\sqrt{n} = 10.32 \pm 3.25 * 0.545/\sqrt{10} = [9.76,\ 10.88]$.

Uppgift 3

a) Genom att stoppa in belopen i uttrycken så får vi $\beta^* = S_{xy}/S_{xx} = 1.051$ och $\alpha^* = \bar{y} - \beta^*\bar{x} = -0.294$. 95% konfidensintervall för de två skattningarna ges av (med skattat s används t-fördelning med 4 frihetsgrader i stället)

$\alpha^* \pm \lambda_{0.025}\sigma\sqrt{\sum_i x_i^2/(nS_{xx})} = -0.294 \pm 1.530 = [-1.824, 1.236]$, och

$\beta^* \pm \lambda_{0.025}\sigma/\sqrt{S_{xx}} = 1.051 \pm 0.2963 = [0.755, 1.347]$.

b) Det gäller att $\beta^* \sim N(\beta, \sigma^2/S_{xx})$. För bevis av detta hänvisas till boken s. 432.

Uppgift 4

a) Korrelationen ges av $\rho^* = r_{xy} = S_{xy}/\sqrt{S_{xx}S_{yy}}$ enligt formelsamlingen. Från givna data får man $S_{xy} = 4444 - 1447 * 106/39 = 511.13$, $S_{xx} = 60933 - 1447^2/39 = 7245.6$ och $S_{yy} = 388 - 106^2/39 = 99.9$. Vi får således att $\rho^* = 0.601$.

b) Vi beräknar $T = \sqrt{n-2}r_{xy}/\sqrt{1 - r_{xy}^2} = 4.57$. Under $H_0 : \rho = 0$ så är denna t-fördelad med $n - 2 = 37$ frihetsgrader. Vi får således ett väldigt signifikant resultat $p << 0.0005$. Vi kan således tydligt förkasta att bonus och tentamenspoäng skulle vara okorrelerade, det tycks råda en tydlig positiv kolleration $\rho > 0$.

c) Att $\rho > 0$ betyder att en student som har många bonuspoäng tenderar även att ha bra poäng på själva tentan. Sensmoralen är således att det är nyttigt att göra alla bonusgivande uppgifter. Utöver att de ger bonus ökar de även chansen att man ska göra bra ifrån sig på själva tentan!

Uppgift 5

a) Det gäller att $E(N_i) = 30 * P(X = i)$, så de ges av $E(N_1) = 15,\ E(N_2) = 7.5,\ E(N_3) = 3.75$ och $E(N_{4+}) = 3.75$.

b) och **c)** Vi får att $Z = 1.67 + 0.83 + 1.35 + 0.02 = 3.87$. Detta ska jämföras med χ^2-fördelning med $4 - 1 = 3$ frihetsgrader. För att förkasta H_0 krävs således att $Z > 7.81$ vilket ju inte gäller. Vi kan således *inte* förkasta nollhypotesen. Så trots att förvånansvärt få omgångar resulterade i Bull's eye direkt (dvs n_1 litet) så avviker inte data signifikant från för-första-gången fördelningen.

Uppgift 6

a) Att testa $\mu_1 - \mu_2$ undersöks lämpligen med $\bar{x} - \bar{y}$. Denna skattning har ju varians $\sigma_1^2/n_1 + \sigma_2^2/n_2$, så den lämpliga referensvariabeln blir

$$R_{\mu_1-\mu_2} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{10} + \frac{s_2^2}{8}}},$$

vilket även är testvariabeln som anges i formelsamlingen. Värdet på denna för våra observerade värden blir $r_{\mu_1-\mu_2} = 2.101$.

b) Från formelsamlingen ser man att denna referensvariabel approximativt följer en t-fördelning under H_0. Antalet frihetsgrader som man ska approximera med har en krånglig formel i formelsamlingen som för våra data blir $f \approx 9.2$ så vi väljer 9 frihestgrader som ligger närmast. Från formelsamlingen ser vi att vi ska förkasta H_0 på 99% nivån om $|R_{\mu_1-\mu_2}| > 3.22$. Eftersom detta inte gäller så förkastas inte H_0 och vi kan därmed *inte* utesluta att de två stickproven har samma vänte värde. Eftersom vi är ganska nära att förkasta finns det dock tecken på att $\mu_1 > \mu_2$.