STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
LÖSNINGAR
16 februari 2023

Lösningar

Tentamen i Statistisk analys, 16 februari 2023

Uppgift 1

a) Falskt

b) Sant

c) Sant

d) Sant

e) Falskt

Uppgift 2

a) Några observationer, framför allt andra observationens värde “efter", är väldigt stora och skulle helt dominera både skattning förändrad försäljning och standardavvikelsen. Därför bättre med icke-parametriskt som betraktar storleksordning i stället för de numeriska värdena (ett skäl emot kan vara att företaget primärt kanske bryr sig om den totala försäljningen som ju är proportionell mot medelvärdet).

b) Den förändrade försäljningen hos respektive individid blir: 1.9, 21.1, -0.2, 2.2, 6.9, 0.6, -0.3, 3.7, 1.4, och 3.7. Ett lämpligt test att använda är teckenrangtestet. Vi summerar rangerna för observationerna för de negativa förändringarna. Det finns två sådana observationer och de är den minsta och tredjen minsta, så vi får $T_- = 1 + 3 = 4$. Enligt tabellen ska vi förkasta nollhypotesen om ingen skillnad (dvs att kursen inte har någon effekt) om

$T_- \leq 8$ eller $T_- \geq 47$ (vi har 2.5% i vardera ände). Vår observation uppfyller detta med råge, så vi förkaster nollhypotesen utan tvekan. Kursen har en tydligt signifikant effekt.

Uppgift 3

a) En lämplig modell här är enkel linjär regression. Om Y anger istjocklek och x latitud ges modellen av $Y_i = \alpha + \beta x_i + \epsilon_i$, där endast ϵ_i är slumpmässig i högerledet, och där dessa slumpfel är oberoende, och (approximativt) normalfördelade med väntevärde 0 och samma varians σ^2.

b) Vi får att $S_{xx} = \sum_i (x_i - \bar{x})^2 = 40$, $S_{xy} = \sum_i x_i y_i - n\bar{x}\bar{y} = 12428 - 5 * 64 * 38 = 268$. Vi får således att $\beta^* = S_{xy}/S_{xx} = 6.7$. Detta är just skattning av förväntad ökad istjocklek per latitud.

c) Vi får $\alpha^* = \bar{y} - \beta^* \bar{x} = 38 - 6.7 * 64 = -390.8$ (en skattning som inte går att tolka - isen kan ju inte ha negativ tjocklek). Punktprediktionen för istjockleken vid en sjö på latitud 63 blir således $\alpha^* + \beta^* * 63 = 31.3$. Standardavvikelsen $s = \sqrt{SSE/(5-2)} = 2.12$. Ett 95% prediktionsintervall för sjöns istjocklek blir således

$$31.3 \pm t_{0.025}(5-2) * s * \sqrt{1 + \frac{1}{5} + \frac{(63-64)^2}{S_{xx}}} = 31.3 \pm 3.182 * 2.12 * \sqrt{1.225},$$

vilket således blir 31.2 ± 7.5. Med andra ord ganska stor osäkerhet, inte minst för att skattningarna baseras på så pass få observationer.

Uppgift 4

a) Ett 95% konfidensintervall för genomsnittlig skärmtid bland svenska ungdomar ges av $\bar{x} \pm t_{0.025}(10-1) * s_x/\sqrt{10} = 2.75 \pm 0.73 = [2.02,\ 3.48]$.

b) Vi skattar skillnaden i skärmtid med $\bar{x} - \bar{y} = 1.35$. Under förutsättning att variationen bland vuxna är densamma som för ungdomar (verkar rimligt med så liknande s) ges ett konfidensintervall för skillnaden av

$$\bar{x} - \bar{y} \pm t_{0.025}(10+12-2) * s_p * \sqrt{1/10 + 1/12} = 1.35 \pm 0.89.$$

Eftersom intervallet är lång ifrån att täcka in nollhypotesens värde om ingen skillnad, ger testet signifikant utfall. Ungdomar har signifikant längre skärmtid än unga vuxna.

Uppgift 5

a) De två observationskoordinaterna får antas vara någorlunda normalfördelade, och olika tentander har oberoende resultat. Visserligen är antal möjliga utfall för bonus bara 0 till 6, men detta får anses någorlunda uppfyllt.

b) Korrelationskoefficienten blir $r_{xy} = S_{xy}/\sqrt{S_{xx}S_{yy}}$. Från uppgiften får vi $S_{xx} = \sum_i x_i^2 - n^{-1}\left(\sum_i x_i\right)^2 = 5901.4$, $S_{yy} = \sum_i y_i^2 - n^{-1}\left(\sum_i y_i\right)^2 = 123.74$ och $S_{xy} = \sum_i x_i y_i - n^{-1}\left(\sum_i x_i\right)\left(\sum_i y_i\right) = 583.61$. Korrelationen ρ skattas således med $r_{xy} = 0.683$. För att testa om $H_0 : \ \rho = 0$ så bildar vi

$$T = \frac{\sqrt{31-2}r_{xy}}{\sqrt{1-r_{xy}^2}} = 5.04.$$

Detta värde ska jämföras med t-fördelningen med 29 frihetsgrader. Detta blir således tydligt signifikant utslag (oavsett av val av signifikansnivå). Slutsatsen är således att det finns et tydligt kausalt samband att den som får hög bonuspoäng tenderar att även få hög poäng på själva tentamen. Detta bör således uppmuntra studenter än mer att se till att arbeta hårt under kursen för att erhålla så många bonuspoäng som möjligt.

c) Enligt formelsamlingen ges ett approximativt konfidensintervall för $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ av

$$\frac{1}{2}\log\left(\frac{1+r_{xy}}{1-r_{xy}}\right) \pm \frac{\lambda_{0.025}}{\sqrt{n-3}}.$$

Eftersom $r_{xy} = 0.683$ så blir första termen 0.8347, och andra termen blir $1.96/\sqrt{28} = 0.37$.

Genom att lösa ut ρ ur $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) = 0.8347 - 0.370$ (och motsvarande med "$+0.370$så får vi att konfidensintervallet för ρ ges av $[\rho^-, \ \rho^+] = [0.43, 0.84]$.

Uppgift 6

a) Båda tätheter har räta linjer. Den första är växande och den andra avtagande.

b) Man beräknar först $E(X) = \int_0^1 x(1+\theta-2\theta x)dx = \frac{1}{2} - \frac{1}{6}\theta$. Momentmetoden erhålls genom att lösa ekationen $\bar{x} = E(X)$ vilket ger $\tilde{\theta}_{MM} = 6\left(\frac{1}{2} - \bar{x}\right)$.

c) Medelvärdet av observationerna är $\bar{x} = 0.655$ så vi får skattningen $\tilde{\theta}_{MM} = -0.93$.