MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 25 oktober 2013 fm

Tillåtna hjälpmedel: Valfri räknedosa utan lagrad information om kursen och utan möjlighet att ansluta till internet direkt eller via wi-fi, Beta samt kursens formel- och tabellsamling.

Examinator: Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Jour: Tommy Norberg.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Obs att ev bonus från duggan endast gäller godkäntnivån.

Svar och lösningar till följande 6 uppgifter skall motiveras om ej annat sägs. Tänk på att sannolikheter bör ges med minst tre korrekta värdesiffror. Avrunda med förstånd för övrigt.

Lösningar publiceras på kurshemsidan.

Uppgifter

1. (a) Antag att händelserna A och B utesluter varandra. Antag vidare att $P(A) = 0.30$ och att $P(B) = 0.55$. Beräkna $P(A \cap B)$. (1 p)

   (b) Antag att händelserna A och B är oberoende. Antag vidare att $P(A) = 0.30$ och att $P(B) = 0.55$. Beräkna $P(A \cup B)$. (1 p)

   (c) Antag att $P(A) = 0.30$, att $P(B) = 0.55$ och att $P(A \cap B) = 0.20$. Beräkna $P(B|A)$. (1 p)

2. En symmetrisk binär överföringslänk har felsannolikheten $p_F = 0.01$. Detta betyder att båda typerna av fel (en etta tolkas som en nolla, samt en nolla tolkas som etta) sker med frekvensen 0.01. Det är vidare känt att 45% av datan som sänds är ettor. Antag nu att man mottagit (a) en etta, resp (b) en nolla. Beräkna i sannolikheten att det var en (a) etta, resp (b) nolla som sändes. Beräkna dessutom (c) hur stor proportion av den mottagna datan som är ettor. (4 p)

3. Antag att cykelolyckor, där minst en av de inblandade parterna behöver sjukhusvård, sker med intensiteten 1.7 per dygn på arbetsdagar och 0.4 per dygn på övriga dagar i Göteborg.

   (a) Hur många sådana olyckor kan man förvänta sig sker under en vecka? (1 p)

   (b) Beräkna sannolikheten att det under två på varandra följande arbetsdagar sker minst 2 olyckor. (2 p)

4. En teknisk utrustning har felbenägenheten $\rho(t) = 0.3t$ för $t > 0$.

   (a) Beräkna förväntad tid tills utrustningen slutar fungera. (2 p)

   (b) Beräkna sannolikheten att utrustningen fungerar högst 1 tidsenhet. (1 p)

   (c) Antag att man i syfte att förbättra tillförlitligheten skaffar ytterligare en sådan utrustning. Hur stor är sannolikheten att systemet bestående av båda utrustningarna parallellkopplade fungerar högst 1 tidsenhet? (1 p)

5. I den här uppgiften ska du analysera data ifrån två på varandra följande partisympatiundersökningar. I den första tillfrågades 1253 potentiella väljare och 129 av dessa svarade Parti A på frågan "Vilket parti skulle du rösta på om det vore val idag?" I den andra ombads 1079 potentiella väljare svara på samma fråga. Då svarade 89 st Parti A. Testa på nivån 5% nollhypotesen att ingen förändring skett i väljaropinionen mot alternativet att Parti A förlorat röster. (4 p)

6. Som du kanske kommer ihåg ifrån föreläsningarna så kan alla bivariate normalvariabler konstrueras ifrån 2 oberoende N$(0, 1)$-variabler. Låt nu Z_1, Z_2 vara oberoende och N$(0, 1)$-fördelade. Sätt

$$X = a_{11}Z_1 + a_{12}Z_2 + c_1$$
$$Y = a_{21}Z_1 + a_{22}Z_2 + c_2$$

där $a_{11} = 3, a_{12} = 4, a_{21} = 16, a_{22} = 12$ och $c_1 = 10, c_2 = 20$. Bestäm parametrarna i den bivariata normalfördelning som styr utfallen av X, Y. (4 p)

Uppgifterna 7 och 8 nedan hör ihop på så sätt att i den sistnämnda ska du använda data ifrån båda.

7. I en del sammanhang, t.ex, genetiska, kan det vara viktigt med stor variation. Antag att man i ett sådant sammanhang observerat standardavvikelsen 9.17 i 10 oberoende lika- och normalfördelade mätningar. Bestäm ett 90% nedåt begränsat konfidensintervall för mätningarnas standardavvikelse. (3 p)

8. (Jfr uppgift 7 ovan.) I syfte att undersöka mätmetoden gjorde man vid ett annat tillfälle ytterligae 5 mätningar. Därvid observerades standardavvikelsen 11.51. Man är rädd att standardavvikelsen kunde förändras från ett mättillfälle till ett annat.

   (a) Anser du att data tyder på att denna rädsla är befogad. (2 p)

   Vid det första tillfället, som redogjordes för i uppgift 7 ovan, uppmättes medelvärdet 17.3, och vid det andra tillfället urhölls medelvärdet 15.4.

   (b) Kan man på basis av denna information hävda att väntevärdena vid de två mättillfällena var olika? (3 p)

Kortfattade lösningar till tentamen MVE090 Matematisk statistik Z, 7.5 hp, den 25/10 2013

1. (a) $P(A \cap B) = 0$, ty $A \cap B = \emptyset$

   (b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.30 + 0.55 - 0.30 \cdot 0.55 = 0.685$

   (c) $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.20}{0.30} = \frac{2}{3} \approx 0.667$

2. Låt $S \in \{0, 1\}$ beteckna signalen som sänds och låt $M \in \{0, 1\}$ beteckna signalen som tas emot. Vi vet att $P(S = 1) = 0.45$, samt att $P(M = 1|S = 0) = P(M = 0|S = 1) = 0.01$. Bayes formel ger nu i (a) resp (b),

$$\begin{aligned} P(S = 1|M = 1) &= \frac{P(S = 1)P(M = 1|S = 1)}{P(S = 1)P(M = 1|S = 1) + P(S = 0)P(M = 1|S = 0)} \\ &= \frac{0.45 \cdot 0.99}{0.45 \cdot 0.99 + 0.55 \cdot 0.01} = 0.988 \\ P(S = 0|M = 0) &= \frac{P(S = 0)P(M = 0|S = 0)}{P(S = 0)P(M = 0|S = 0) + P)S = 1)P(M = 0|S = 1)} \\ &= \frac{0.55 \cdot 0.99}{0.55 \cdot 0.99 + 0.45 \cdot 0.01} = 0.992 \end{aligned}$$

och enl lagen om total sannolikhet får vi (c)

$$\begin{aligned} P(S = 1) &= P(S = 1)P(M = 1|S = 1) + P(S = 0)P(M = 1|S = 0) \\ &= 0.45 \cdot 0.99 + 0.55 \cdot 0.01 = 0.451 \end{aligned}$$

3. (a) Under en vecka kan man förvänta sig $5 \cdot 1.7 + 2 \cdot 0.4 = 9.3$ olyckor

   (b) Antalet olyckor under två på varandra följande arbetsdagar följer en Poissonfördelning med parameter $k = 2 \cdot 1.7 = 3.4$, så den sökta sannolikheten är

$$p = 1 - e^{-3.4}\left(\frac{3.4^0}{0!} + \frac{3.4^1}{1!}\right) = 1 - 4.4e^{-3.4} = 0.8532$$

4. $\rho(t) = 0.3t \Rightarrow \int_0^t \rho(s)\, ds = 0.15t^2$, så $R(t) = e^{-0.15t^2}$ och $f(t) = 0.3te^{-0.15t^2}$. Detta är en Weibulltäthet med $\alpha = 0.15$ och $\beta = 2$.

   (a) $\mu = \alpha^{-1/\beta}\Gamma(1 + 1/\beta) = 0.15^{-1/2} \cdot 0.5 \cdot \sqrt{\pi} = 2.288$

   (b) $F(1) = 1 - R(1) = 1 - 0.0.8607 = 0.139$

   (c) $(1 - R(1))^2 = F(1)^2 = 0.0194$

5. Vi har $f_1 = 129$, $n_1 = 1253 \Rightarrow \hat{p}_1 = \frac{129}{1253} = 0.1030$ samt $f_2 = 89$, $n_2 = 1079 \Rightarrow \hat{p}_2 = \frac{89}{1079} = 0.0825$. Under $H_0 : p_1 = p_2 = p$ skattar vi p med $\hat{p} = \frac{f_1 + f_2}{n_1 + n_2} = \frac{218}{2332} = 0.0935$. Teststatistikans värde är

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}} = \frac{0.1030 - 0.0825}{\sqrt{0.0935 \cdot (1 - 0.0935)(1/1253 + 1/1079)}} = 1.696$$

P-värdet är $P(Z > z) = P(Z > 1.696) = 1 - \Phi(1.70) = 0.0446$. Nollhypotesen $H_0 : p_1 = p_2$ kan således förkastas till förmån för alternativet $H_1 : p_1 > p_2$ på nivån 5%.

6. Parametrarna som ska bestämmas är variablernas väntevärden, standardavvikelser och deras korrelation. Vi har $EX = c_1 = 10$, $EY = c_2 = 20$, $\mathrm{Var}(X) = a_{11}^2 + a_{12}^2 = 3^2 + 4^2 = 25 = 5^2$, $\mathrm{Var}(Y) = a_{21}^2 + a_{22}^2 = 16^2 + 12^2 = 400 = 20^2$, samt $\mathrm{Kov}(X,Y) = E(a_{11}Z_1 + a_{12}Z_2)(a_{21}Z_1 + a_{22}Z_2) = a_{11}a_{21}EZ_1^2 + a_{11}a_{22}EZ_1Z_2 + a_{12}a_{21}EZ_1Z_2 + a_{12}a_{22}EZ_2^2 = a_{11}a_{21} + a_{12}a_{22} = 3 \cdot 16 + 4 \cdot 12 = 96$. Detta ger korrelationen $\rho = \frac{96}{5 \cdot 20} = 0.96$.

7. Givet är $s_1 = 9.17$, $n_1 = 10$ och att observationerna i stickprovet är $\mathrm{N}(\mu, \sigma)$. Undre gräns för σ med 90% konfidens är

$$\sqrt{\frac{n-1}{\chi^2_{0.10}}} \cdot s = \sqrt{\frac{9}{14.6837}} \cdot 9.17 = 7.18$$

8. (a) Givet är $s_1 = 9.17$, $n_1 = 10$, samt $s_2 = 11.51$, $n_2 = 5$. Stickproven är oberoende och skattar σ_1 resp σ_2. Under $H_0 : \sigma_1 = \sigma_2$ är teststatistikan $s_2^2/s_1^2 \sim F(4,9)$. Observerat värde $11.51^2/9.17^2 = 1.576$ är inte större än $f_{0.10}(4,9) = 2.693$. Slutsats: Inte ens på 20%-nivån kan vi förkasta H_0. Data ger oss alltså ingen anledning att ens misstänka att $\sigma_1 \neq \sigma_2$.

   (b) Vi bildar den sammanvägda variansen

$$s_P^2 = \frac{9 \cdot 9.17^2 + 4 \cdot 11.51^2}{9+4} = 98.9785 = 9.949^2$$

   Observerat värde på teststatistikan, som är $t(13)$-fördelad, är

$$t = \frac{1.9}{9.949\sqrt{1/10 + 1/5}} = 0.349$$

   Jfr $t_{0.1}(13) = 1.350$. Slutsats: Inte ens med max 20% risk att ha fel kan vi hävda att väntevärdena var olika.