MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 24 oktober 2012 em

Tillåtna hjälpmedel: Valfri räknedosa utan lagrad information om kursen, 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 8 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. Betrakta tre händelser A, B och C. Antag att

$$P(A) = 0.7, \quad P(B) = 0.6, \quad P(C) = 0.5,$$
$$P(A \cap B) = 0.4, \quad P(A \cap C) = 0.3, \quad P(B \cap C) = 0.2,$$
$$P(A \cap B \cap C) = 0.1$$

   Beräkna $P(A \cup B \cup C)$. (3 p)

2. I staden A regnar det i snitt var 3:e förmiddag. G gillar att cykla och tar gärna cykeln till jobbet även om det regnar. Antag att G cyklar fyra av fem regndagar och alla övriga dagar.

   (a) Hur vanligt är det att G cyklar till jobbet? (2 p)

   (b) På kvällen får du reda på att idag cyklade G till jobbet. Hur stor är då den betingade sannolikheten att det inte regnade i förmiddags? (2 p)

3. Var och en av skollärare M:s 20 sjundeklasselever singlar en slant 100 gånger. Ungefär hur många förväntas få mellan 40 och 60 klave? Här går det bra med överslagsberäkningar grundade på tumregler. (3 p)

4. För att en applikation ska fungera tillfredsställande, måste tre delsystem, kallade A, B och C, fungera. Delsystem A har tillförlitligheten (funktionssannolikheten) 0.990. Delsystem B består av två enheter och det räcker att en av dessa fungerar tillfredsställande. Dessa två enheter har båda tillförlitligheten 0.900. Delsystem C består av tre enheter, alla med tillförlitligheten 0.941, och minst två av de tre enheterna måste fungera.

   (a) Hur stor är enhet B:s tillförlitlighet? (1 p)

   (b) Hur stor är enhet C:s tillförlitlighet? (2 p)

   (c) Hur stor är applikationens tillförlitlighet? (1 p)

   Om du i (b) inte kan beräkna C:s tillförlitlighet, använd värdet 0.990 i (c). Vi antar att alla delsystem och deras enheter fungerar oberoende av varandra.

5. Låt funktionen

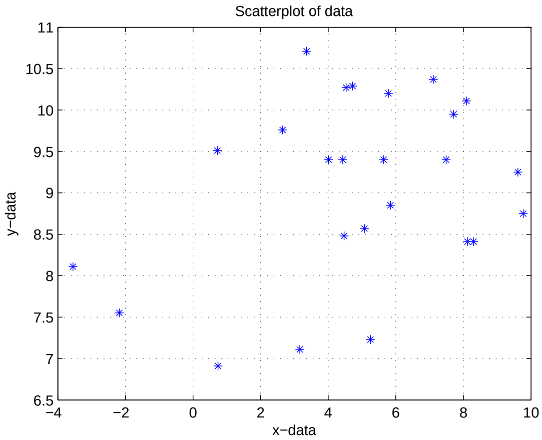
$$f_{XY}(x,y) = \frac{x+2y}{c} \quad \text{för} \quad 0 \le x < y \le 5$$

   vara täthet för en två-dimensionell stokastisk variabel X, Y. Beräkna, för $0 \le y \le 5$, det betingade väntevärdet av X givet $Y = y$. (4 p)

6. Vid ett tillfälle mättes resistansen i 15 slumpmässigt utvalda elektriska motstånd. Därvid erhölls medelvärdet 2.2 ohm och standardavvikelsen 0.73. Bestäm ett uppåt begränsat konfidensintervall för standardavikelsen med konfidens 95%. Antag normalfördelade observationer. (3 p)

7. I den här uppgiften ska du analysera utfallet av två på varandra följande partisympatienkäter. I den första tillfrågades 1263 slumpmässigt utvalda väljare, varav 180 svarade att de skulle rösta på partiet B om det vore val idag. Innan den andra som gjordes en månad senare gjorde partiet B ett utspel som man hoppades skulle medföra ökad popularitet. Men det fanns också de som menade att det lika gärna kunde slå åt det andra hållet. I den andra undersökningen tillfrågades 1702 slumpmässigt utvalda väljare, varav 291 svarade att de skulle rösta på partiet B om det vore val idag, som alltså var ungefär en månad senare. Punkt- och intervallskatta förändringen av partiets väljarandel. Önskad konfidensgrad: 0.95. (4 p)

8. Här nedan ser du en s.k scatterplot av ett bivariat stickprov (x_i, y_i), $i = 1, \dots, n$:



Du förväntas lösa följande två deluppgifter.

(a) Skatta regressionslinjen $y = \beta_0 + \beta_1 x$. Vi antar att data är oberoende observationer $Y_1, \dots, Y_n$ på nivåerna $x_1, \dots, x_n$ och sådana att $Y_i \sim \mathrm{N}(y_i, \sigma)$, där $y_i = \beta_0 + \beta_1 x_i$. (2 p)

(b) Skatta korrelationen ρ och testa på nivån 0.05 nollhypotesen $H_0 : \rho = 0$ mot alternativet $H_1 : \rho > 0$. Vi antar här att data är oberoende observationer (x_i, y_i) på en bivariat normalvariabel X, Y med korrelationen ρ. (3 p)

Data kan sammanfattas så här: $\sum x = 120.84$, $\sum y = 226.40$, $\sum x^2 = 851.9686$, $\sum y^2 = 2078.5970$, $\sum xy = 1123.5385$. Antalet observationer är $n = 25$.

1. Lättast är nog att rita ett Venn-diagram och räkna utgående ifrån det. Annars härleder man (t.ex med hjälp av Venn-diagrammet) eller kommer ihåg att $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) = \cdots = 1$.

2. Låt R vara händelsen att det regnbar och C vara händelsen att G cyklar till jobbet. Vi får reda på att $P(R) = 1/3$, $P(C|R) = 4/5$ och $P(C|R') = 1$. (a) Sökt är $P(C) = P(R)P(C|R) + P(R')P(C|R') = 14/15 = 0.933$ (lagen om total sannolikhet). G cyklar alltså i snitt 14 dagar av 15. (b) Sökt är $P(R'|C) = P(R')P(C|R')/P(C) = 5/7 = 0.714$ (Bayes formel).

3. Antalet klave är $\text{Bin}(n = 100, p = 0.5)$. Förväntat värde är $np = 50$ och standardavvikelsen är $\sqrt{np(1-p)} = 5$. Med hjälp av en normalapproximation av binomial kan sannolikheten för 40–60 klave, jämföras med sannolikhen att en normalvariabels utfall hamnar i intervallet $\mu \pm 2\sigma$ och vi vet att denna sannolikhet är ungefär 0.95. Vi kan därför se varje elevs 100 slantsinglingar som ett försök i vilket sannolikheten att det lyckas är ungefär 0.95. De 20 försöken är naturligtvis oberoende, så vi kan förvänta oss att ca $20 \cdot 0.95 = 19$ lyckas. Svaret på uppgiften är alltså 19.

4. (a) $p_B = 1 - (1 - 0.9)(1 - 0.9) = 0.99$ (b) $p_C = \binom{3}{2} 0.941^2 (1 - 0.941) + \binom{3}{3} 0.941^3 = 0.990$ (c) $p = p_A p_B p_C = 0.990^3 = 0.970$.

5. $f_Y(y) = \int_0^y \frac{x+2y}{c}\, dx = \frac{5y^2}{2c}$ för $0 \le y \le 5 \Rightarrow f_{X|y}(x) = \frac{(x+2y)/c}{5y^2/(2c)} = \frac{2}{5}\frac{x+2y}{y^2}$ för $0 \le x < y \Rightarrow E[X|Y = y] = \int_0^y x \frac{2}{5} \frac{x+2y}{y^2}\, dx = 8y/15$ för $0 \le y \le 5$.

6. Utgå ifrån $(n-1)s^2/\sigma^2 \sim \chi^2(n-1)$. Vi har alltså $n-1 = 14$ frihetsgrader och får ur lämplig χ^2-tabell 0.95-kvantilen 6.571. Obs att 0.95-kvantilen definieras av att $P(\chi^2(n-1) > x_{0.95}) = 0.95$. Vi har alltså att $P(6.571 \le 14s^2/\sigma^2) = 0.95$. Omvandla olikheten, sätt in uppmätt standardavvikelse och konstatera att $\sigma^2 \le 14 \cdot 0.73^2/6.571 = 1.1354$, d.v.s $\sigma \le \sqrt{1.1354} = 1.066$ gäller med 95% konfidens.

7. Punkskattningen av förändringen är $\widehat{p_2 - p_1} = \hat{p}_2 - \hat{p}_1 = 0.1425 - 0.1710 = 0.0285$ och den statistiska felmarginalen (95% konfidens) är $\pm 1.96\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} = \pm 0.0263$. Data tyder med minst 95% konfidens på att det skett en förändring. Alternativt kan man beräkna en undre gräns för $p_2 - p_1$ med konfidens 95% enligt $0.0285 - 1.645\sqrt{0.1425(1-0.1425)/1263 + 0.1710(1-0.1710)/1702} = 0.0285 - 0.0221 = 0.0064$. Data tyder då med minst 95% konfidens på att väljarandelen ökat.

8. Beräkna först $S_{xx} = \sum x^2 - (\sum x)^2/n = 267.8764$, $S_{yy} = \sum y^2 - (\sum y)^2/n = 28.3186$ och $S_{xy} = \sum xy - (\sum x)(\sum y)/n = 29.2115$. (a) Fortsätt sedan med $b_1 = 29.2115/267.8764 = 0.1090$ och $b_0 = 226.4/25 - b_1 120.84/25 = 8.5289$. Regressionslinjen är således $y = 8.529 + 0.109x$. (b) Här får vi estimatet $r = 29.2115/\sqrt{267.8764 \cdot 28.3186} = 0.3354$ av korrelationen ρ. Obs värde på teststatistikan $T = R\sqrt{n-2}/\sqrt{1-R^2} \sim t(n-2)$ är $t = 1.707$. J.f.r med $t_{0.05}(23) = 1.714$. Vi kan uppenbarligen inte förkasta H_0 på nivån 5%.