MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 19 oktober 2009 em

Tillåtna hjälpmedel är räknedosa utan lagrad information om kursen, Beta samt kursens formel- och tabellsamling.

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

Övningsledare är Anna Rudvik, ankn 5338 eller 0730 57 96 26.

Jour är Anna.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan.

Svar och lösningar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. Kvalitetskontroll. Komponenter inköpes till en produktionslinje n åt gången. Antag att Y av dessa är defekta. Vid ankomst väljes $1 \leq k \leq n$ st slumpmässigt ut för kontroll. Låt X vara antalet defekta i urvalet.

   (a) Bestäm den betingade tätheten $f_{X|y}(x)$ för X givet att $Y = y$. (3 p)

   Antag att Y:s täthet är

   $$f_Y(y) = \frac{1}{n+1} \quad \text{för} \quad 0 \leq y \leq n$$

   (b) Bestäm den betingade tätheten för Y givet $X = x$, alltså $f_{Y|x}(y)$. Förenkla så långt som möjligt. (2 p)

   Kommentar: Medelst maximering av $f_{Y|x}(y)$ m.a.p y för ett givet fixt x, får man trolighetsskattningen av Y (antalet defekta enheter).

2. Vid tillverkning av resistorer sorterar man ut till försäljning de som ligger inom $\pm 10\%$ från nominell resistans. Övriga kasseras. Antag att avvikelsen ifrån nominell resistans är normalfördelad med standardavvikelse ca 5% av nominellt värde. Av 100 tillverkade resistorer, ungefär hur många kommer att kasseras? (3 p)

3. Felbenägenheten för en av företaget E:s produkter ges av uttrycket

   $$z(x) = 0.0862x^{-1.3} \quad \text{för} \quad x > 0$$

   där tiden x räknas i 10^3 timmar. Ungefär hur många timmar överlever 90% av de tillverkade enheterna? (4 p)

4. Låt Z_1, Z_2 vara två oberoende normalfördelade variabler med väntevärde 0 och standardavvikelse 1. Sätt

   $$X = 10 + 2.5Z_1$$
   $$Y = 15 + 1.2Z_1 + 1.6Z_2$$

   Man kan visa att X, Y har en s.k bivariat normalfördelning.

(a) Vilka värden har parametrarna i denna? (3 p)

(b) Bestäm den betingade fördelningen för Y givet $X = x$. (2 p)

5. Man ville undersöka huruvida ett ev systematiskt fel ("bias") föreligger i tillverkningen av 100 Ohms-resistorer. J.f.r uppgift 2. Man mätte därför med ett precisionsmätinstrument upp $n = 25$ st från tillverkningen slumpvis valda 100 Ohms-resistorer. Därvid erhölls värdena

99.21	93.04	102.00	102.81	95.64
107.32	107.32	101.18	103.01	102.24
100.44	105.00	98.43	112.29	100.69
101.94	106.70	101.67	100.89	97.21
102.84	94.69	104.94	109.49	97.91

(Räknehjälp: $\sum x = 2\,548.90$, $\sum x^2 = 26\,0376.8972$; obs att om du ej klarar av att beräkna medelvärde och standardavvikelse så kan du mot 1 p totalt avdrag använda värdena $\bar{x} = 102.00$, $s = 4.50$ här och i uppgift 7.) Undersök huruvida ett systematiskt fel föreligger. Vilken typ av fel görs möjligen? Ange, om möjligt, ungefär hur stor risken maximalt kan vara att ett sådant fel görs. (4 p)

6. Ett annorlunda och kanske lite enklare sätt att avgöra huruvida bias föreligger kan göras genom att endast analysera huruvida de uppmätta resistanserna ligger över eller under nominellt värde. I datan i uppgift 5 är 7 observationer under det nominella värdet 100 Ohm. Punkt- och intervallskatta sannolikheten p att det uppmätta värdet ligger under det nominella. Går det att ur denna analys dra någon slutsats om huruvida bias föreligger? Gör i så fall det och ange, om möjligt, ungefär hur stor risken maximalt kan vara att ett sådant fel görs. (3 p)

7. I samband med att förekomsten av ev bias studerades (j.f.r uppgifterna 2 och 5) kan vi passa på att undersöka uppgiften att standardavvikelsen är ca 5% av nominellt värde. Basera dina svar på följande frågor på en relevant analys av datan i uppgift 5. Stämmer det att standardavvikelsen är ca 5? Svara ja eller nej. Vilken typ av fel gör du då möjligen? Ange, om möjligt, ungefär hur stor risken maximalt kan vara att du gör ett sådant fel. (3 p)

8. Låt X, Y vara som i uppgift 4 bivariat normalfördelade. Vid ett tillfälle gjordes $n = 10$ oberoende observationer av X, Y. Härvid erhölls $\sum x = 98.60$, $\sum y = 150.66$, $\sum x^2 = 1\,022.4263$, $\sum y^2 = 2\,293.4799$ och $\sum xy = 1\,497.7808$. Testa på nivån 10% huruvida X och Y är korrelerade. (3 p)

Lycka till med lösandet av uppgifterna!

Lösningar till Mat stat Z den 19/10-2009

1. (a) Ur de y defekta ska vi dra x st och de återstående $k-x$ ska dras från de återstående $n-y$ komponenterna. Enl multiplikationsprincipen kan detta ske på $\binom{y}{x}\binom{n-y}{k-x}$ olika sätt. Totala antalet sätt att dra k komponenter ur en mängd med n st är $\binom{n}{k}$. Eftersom alla utfall är lika sannolika, gäller

$$f_{X|y}(x) = \frac{\binom{y}{x}\binom{n-y}{k-x}}{\binom{n}{k}}$$

Här är $0 \leq x \leq \min(y, k)$.

(b) Enl definitionen av betingning gäller att

$$f_{X,Y}(x,y) = f_Y(y) f_{X|y}(x) = \frac{1}{n+1} \frac{\binom{y}{x}\binom{n-y}{k-x}}{\binom{n}{k}}$$

Samma definition ger även att

$$f_{Y|x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

där

$$f_X(x) = \sum_{y=x}^{n} f_{X,Y}(x,y) = \sum_{y=x}^{n} \frac{1}{n+1} \frac{\binom{y}{x}\binom{n-y}{k-x}}{\binom{n}{k}}$$

Härur följer

$$f_{Y|x}(y) = \frac{\binom{y}{x}\binom{n-y}{k-x}}{\sum_{y=x}^{n} \binom{y}{x}\binom{n-y}{k-x}}$$

för $x \leq y \leq n$.

2. Ca 95% av alla observationer av normalvariabler landar inom ± 2 standardavvikelser ifrån väntevärdet. Härur följer att ca 5% av alla resistorer kommer att förkastas. Svaret på frågan är därför 5.

3. Obs feltryck i tesen. Det ska vara $z(x) = 0.0862x^{-0.6}$. Den smarte använder faktumet $R(x) = e^{-\int_0^x z(y)\,dy}$ och får

$$R(x) = e^{-0.2155x^{0.4}}$$

Låt x vara den sökta tiden. Då följer ur $R(x) = 0.9$ att

$$x = \left(\frac{-\ln 0.9}{0.2155}\right)^{10/4} = 0.1671$$

Svar: 167.1 timmar

4. (a) Man ser direkt att $\mu_x = 10$ och $\mu_y = 15$, samt att $\sigma_x = 2.5$ och $\sigma_y = \sqrt{1.2^2 + 1.6^2} = 2$. Vidare, $\mathrm{Cov}[X,Y] = E[2.5Z_1(1.2Z_1 + 1.6Z_2)] = 2.5 \cdot 1.2 = 3$ och det följer att korrelationen är $\rho = \frac{3}{2 \cdot 2.5} = 0.6$.

(b) $X = x \Leftrightarrow Z_1 = \frac{x-10}{2.5} \Rightarrow Y = 15 + 1.2 \cdot \frac{x-10}{2.5} + 1.6Z_2 = 10.2 + 0.48x + 1.6Z_2$. Således gäller att $Y|X = x \sim \mathrm{N}(10.2 + 0.48x,\ 1.6)$.

På tentan är det naturligtvis viktigt att de icke självklara beräkningsstegen motiveras. Jag har här avstått ifrån detta av pedagogiska skäl.

5. Ur data fås $\bar{x} = 101.956$, $s = \sqrt{20.8854} = 4.5701$ och $n = 25$. Ett 95% konfidensintervall är $\mu = \bar{x} \pm t_{0.025}s/\sqrt{n} = 101.956 \pm 2.0639 \cdot 4.5701/\sqrt{25} = 101.956 \pm 1.886$. Obs att värdet $\mu = 100$ ej tillhör intervallet. Så med minst 95% konfidens kan vi hävda att det föreligger ett systematiskt fel. Här görs ev ett s.k typ-I-fel och risken att vi gjort ett sådant fel är ≤ 0.05. (De som räknar med värdena $\bar{x} = 102$ och $s = 4.5$ gör samma analys fast med aningen annorlunda siffror.)

6. Är inget sagt, väljes konfidensen 95% (detta gäller även ovanstående uppgift). Så, enl formelsamling,

$$p = \frac{7}{25} \pm 2.064\sqrt{\frac{\frac{7}{25} \cdot \frac{18}{25}}{24}} = 0.28 \pm 0.189$$

Övre gränsen i detta intervall är $0.469 < 0.5$. Är det sannt att $p \leq 0.469$ så är det även sannt att $p < 0.5$, vilket innebär att bias föreligger. Vi kan alltså påstå att bias föreligger och vi gör så med max risk att ha fel ungefär 5%. Den minnesgode vet från föreläsning att vi helst ska ha frekvensen $10 \leq f \leq n - 10 = 15$ för att normalapproximation som görs ska vara ok. Analysen är alltså aningen tveksam. Å andra sidan är 0.469 en bit ifrån 0.5, så man kan våga tro att risken att vi har fel då vi påstår att systematiskt fel föreligger inte är större än ca 5%.

7. Här ska vi se om det går att förkasta nollhypotesen $H_0 : \sigma = 5$ mot alternativet $H_1 : \sigma \neq 5$. De två kvantiler som är aktuella i $\chi^2(24)$-fördelningen är de som svarar mot sannolikheterna 0.025 och 0.975. De är 12.401 och 39.364. Test-statistikan har värdet

$$x^2 = \frac{24 \cdot 20.885}{5^2} = 20.050$$

Detta värde ligger klart innanför gränserna 12.401 och 39.364, så det går inte att förkasta $H_0 : \sigma = 5$ på nivån 5% och förmodligen inte heller på 10%-nivån. Ett annat sätt att komma till denna slutsats är att räkna ut 95%-intervallet $3.57 \leq \sigma \leq 6.36$. Det finns alltså inget i datan som tyder på att $H_1 : \sigma \neq 5$ är sann. Möjligen gör vi ett s.k typ-II-fel (d.v.s en falsk icke-detektion). Det går inte att uttala sig om risken att vi gjort ett sådant fel. Det man generellt kan säga är att den minskar med stickprovsstorleken och att vi faktiskt har ganska många observationer, så man kan kanske hoppas att den felrisken inte är så stor.

8. Räkna ut $S_{xx} = \sum x^2 - (\sum x)/n = 50.2303$ och analogt $S_{yy} = 23.6363$ samt $S_{xy} = 12.2732$, och sedan skattningen $r = S_{xy}/\sqrt{S_{xx}S_{yy}} = 0.356$. Teststatistikans värde $t = r\sqrt{n-2}/\sqrt{1-r^2} = 1.078$ ska jämföras med $t_{0.05}(8) = 1.860$ och vi ser att hypotesen $H_0 : \rho = 0$ inte kan förkastas ens på nivån 10%. Det fel vi möjligen gör är typ-II (en falsk icke-detektion) och det går inte att uttala sig om risken att ett sådant fel gjorts.