MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 20 oktober em V

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

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

Jour är Anna Rudvik, ankn 5338.

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. Rättningsprotokoll anslås ej.

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

Uppgifter

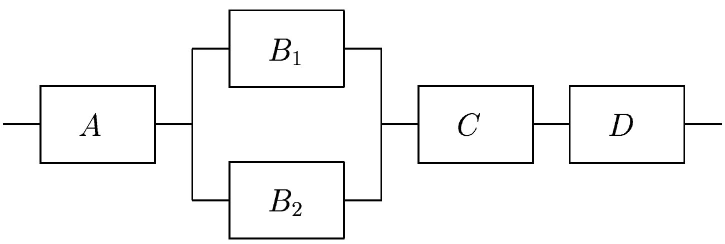
1. I s.k ”screening” gäller att samtliga individer i en viss population undersöks m.a.p någon egenskap. Antag att populationen består av alla kvinnor i Sverige födda ett visst år och att man medelst mammografi vill hitta alla som har bröstcancer. Antag att proportionen kvinnor i åldersgruppen som har cancer är ca 0.0027. De flesta undersökningar av den här typen är belagda med fel och vi ska anta att falsklarmssannolikheten är ca 5% och att upptäcktssannolikheten är ca 97.5%. Detta innebär att 5% av alla testade friska kvinnor kommer att kallas till nya undersökningar p.g.a att testet felaktigt detekterar cancer, samt att man endast hittar 97.5% av alla sjuka testade. Antag nu att man testat en kvinna och att testet har indikerat att denna kvinna har bröstcancer. Hur stor är sannolikheten att hon verkligen har bröstcancer? Reflektera gärna över resultatet. (4 p)

2. Som bekant har Geo(p)-fördelningen tätheten

$$f(x) = (1-p)^{x-1}p \quad \text{för} \quad x = 1, 2, \ldots$$

Beräkna momentgenererande funktion och visa med hjälp av denne att fördelningen har väntevärde $\mu = 1/p$. (3 p)

3. En teknisk konstruktion är m.a.p tillförlitlighet uppdelat i ett antal komponenter enligt följande block-diagram:



Komponenterna A, C och D har exponentialfördelade livslängder med väntevärden $\beta_A = 10$, $\beta_C = 20$ resp $\beta_D = 15$ tidsenheter (t.e). För komponenterna B_1

och B_2 gäller att felbenägenheten (”hazard rate”) är $\rho_{B_1}(t) = \rho_{B_2}(t) = t/20$. Beräkna sannolikheten att konstruktionen fungerar i minst 10 t.e. Antag därvid att komponenternas livslängder är oberoende. (4 p)

4. Låt $X, Y \sim \mathrm{N}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. Visa att

$$Y|X = x \sim \mathrm{N}\left(\mu_y + \rho\,\frac{\sigma_y}{\sigma_x}\,(x - \mu_x)\,,\ \sigma_y\sqrt{1-\rho^2}\right) \qquad (4\text{ p})$$

Ledning: Använd att

$$\begin{aligned} X &= \mu_x + \sigma_x Z_1 \\ Y &= \mu_y + \sigma_y\left(\rho Z_1 + \sqrt{1-\rho^2}\,Z_2\right) \end{aligned}$$

för oberoende $Z_1, Z_2 \sim \mathrm{N}(0,1)$.

5. I många städer mäts ozonhalten regelbundet. Antag att man under en kort tidsperiod erhållit följande mätvärden: 0.425, 0.416, 0.326, 0.375, 0.339 ppm. Punkt- och intervallskatta med konfidensgraden 95% medelhalten. Antag normalfördelade observationer. (3 p)

6. För datamängden i föregående uppgift, beräkna ett uppåt begränsat konfidensintervall för standardavvikelsen med konfidensgrad 99%. (4 p)

7. I staden som refereras till i uppgift 5 ovan, begränsades biltrafiken i syfte att få ned de väl höga ozonhalterna. Man bedömmer att inte bara halten ska minska utan också variationerna i den. Ett tag efter det att trafiken hade minskat mättes ozonhalten ånyo. Därvid erhölls i 10 mätningar $\sum x = 2.983$ ppm och $\sum x^2 = 0.8955$. Går det att visa statistiskt att variationen i halten har blivit mindre?(4 p)

8. Ståls hållfasthet beror av dess kolhalt. I 5 prover uppmättes halterna 1.56, 0.85, 3.01, 1.96, 2.20 och motsvarande hållfastheter var 22.00, 22.54, 27.93, 22.40, 21.15. (Enheterna är oväsentliga.) Går det på basis av denna lilla mätning att säkerställa statistiskt att hållfastheten och kolhalten är beroende? Antag normalfördelade data. (Räknehjälp: $\sum x = 9.58$, $\sum y = 116.02$, $\sum x^2 = 20.8978$, $\sum xy = 227.9823$, $\sum y^2 = 2721.2190$.) (4 p)

Data e dyl i uppgifterna ovan är påhittade och ev likheter med verkliga förhållanden är därför tillfälliga.

Lycka till

Lösningar till Mat stat Z den 20/10-2008

1. Bayes formel ger

$$\frac{0.0027 \cdot 0.975}{0.0027 \cdot 0.975 + (1 - 0.0027) \cdot 0.05} \approx 0.0501$$

Trots att testet med hög säkerhet detekterar cancer är alltså sannolikheten att kvinnan under test verkligen har cancer endast ca 5%. Tänkvärt eller...

2. Låt $q = 1 - p$. Vi får

$$m(t) = \sum_{x=1}^{\infty} e^{tx} q^{x-1} p = \frac{p}{q} \sum_{x=1}^{\infty} \left(qe^t\right)^x = \frac{p}{q}\left(\frac{1}{1-qe^t} - 1\right) = \frac{pe^t}{1-qe^t}$$

för t, sådana att $qe^t < 1$, d.v.s $t < -\ln q$ (den geometriska serien konvergerar ju då kvoten $qe^t < 1$). Notera nu att

$$m'(t) = \frac{pe^t(-qe^t) - pe^t(1-qe^t)}{(1-qe^t)^2} = \frac{pe^t}{(1-qe^t)^2}$$

vilket ger

$$\mu = m'(0) = \frac{p}{(1-q)^2} = \frac{1}{p}$$

3. För komponenterna A, C och D gäller att deras resp tillförlitligheter efter 10 t.e är

$$p_A = e^{-10/10} = e^{-1} = 0.3679 \quad p_C = e^{-10/20} = e^{-1/2} = 0.6065 \quad p_D = e^{-10/15} = e^{-2/3} = 0.5134$$

medan vi får för båda B-komponenterna tillförlitligheten

$$p_{B_i} = e^{\int_0^{10} (t/20)\,dt} = e^{-100/40} = e^{-5/2} = 0.0821$$

Dessa är ju parallell-kopplade, så vi erhåller den resulterande tillförlitligheten

$$p_B = 1 - (1 - e^{-5/2})^2 = e^{-5/2}\left(2 - e^{-5/2}\right) = 0.1574$$

för B-delen. Hela konstruktionens tillförlitlighet efter 10 t.e är

$$p_A\, p_B\, p_C\, p_D = e^{-1} \cdot e^{-5/2}\left(2 - e^{-5/2}\right) \cdot e^{-1/2} \cdot e^{-2/3} \approx 0.0180$$

4. Då $X = x$ måste $Z_1 = (x - \mu_x)/\sigma_x$. Följaktligen,

$$\begin{aligned} Y|x &= \mu_y + \sigma_y \left(\rho \frac{x - \mu_x}{\sigma_x} + \sqrt{1-\rho^2} Z_2\right) \\ &= \mu_y + \rho\, \frac{\sigma_y}{\sigma_x}\,(x - \mu_x) + \sigma_y \sqrt{1-\rho^2} Z_2 \end{aligned}$$

Vi inser av detta att den betingade fördelningen för Y givet $X = x$ är normal med v.v $\mu_y + \rho\, \frac{\sigma_y}{\sigma_x}\,(x - \mu_x)$ och standardavvikelse $\sigma_y\sqrt{1-\rho^2}$, v.s.v.

5. Jag erhåller $\sum x = 1.881$ och $\sum x^2 = 0.7155$, och kan sedan beräkna

$$\bar{x} = \frac{1.881}{5} = 0.376 \quad \text{och} \quad s^2 = \frac{1}{4}\left(0.7155 - \frac{1.881^2}{5}\right) = 0.001968 = 0.04436^2$$

Så (t-kvantil, $n-1=4$ frihetsgrader, $\alpha/2 = 0.025$),

$$\mu = 0.376 \pm 2.77645 \cdot 0.04436/\sqrt{5} = 0.376 \pm 0.055$$

6. Händelsen $(n-1)s^2/\sigma^2 \geq 0.2971 = \chi^2_{0.99}(4)$ inträffar mesd sannolikheten 0.99. Så,

$$\sigma \leq \sqrt{\frac{(n-1)s^2}{0.2971}} = 3.669s \approx 0.163$$

Intervallet $[0,\ 0.163]$ för σ har alltså konfidensen 99%.

7. Variansen är

$$s^2 = \frac{1}{9}\left(0.8955 - \frac{2.983^2}{10}\right) = 0.0006301$$

Varianskvoten, so i detta fall $= 0.001968/0.0006301 = 3.123$, är $F(4,9)$-fördelad och 0.1-kvantilen i denna fördelning är $2.693 < 3.123$, så med 10% risk att ha fel kan vi hävda att variansen har minskat. Vi kan däremot inte hävda detta med 5% risk eftersom 0.05-kvantilen är $3.633 > 3.123$.

8. ρ skattas av

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

där

$$\begin{aligned}
S_{xy} &= \textstyle\sum xy - \left(\sum x\right)\left(\sum y\right)/n = 227.9823 - 9.58 \cdot 116.02/5 = 5.6880 \\
S_{xx} &= \textstyle\sum xx - \left(\sum x\right)\left(\sum x\right)/n = 20.8978 - 9.58 \cdot 9.58/5 = 2.5425 \\
S_{yy} &= \textstyle\sum yy - \left(\sum y\right)\left(\sum y\right)/n = 2721.2190 - 116.02 \cdot 116.02/5 = 29.0909
\end{aligned}$$

Således är $r = 5.688/\sqrt{2.5425 \cdot 29.0909} = 0.661$. Vidare gäller att

$$\frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}} \sim t(n-2) \quad \text{under} \quad H_0 : \rho = 0$$

Vi jämför observerat värde $r\sqrt{n-2}/\sqrt{1-r^2} = 0.661\sqrt{3}/\sqrt{1-0.661^2} = 1.526$ med kvantiler i $t(3)$-fördelningen och ser att det inte går att förkasta $H_0 : \rho = 0$ ens på 20%-nivån eftersom $t_{0.10}(3) = 1.638 > 1.526$.