LINKÖPINGS UNIVERSITET

Matematiska institutionen

16 augusti 2021, kl. 08.00-12.00
Examinator: Jörg-Uwe Löbus

Tillåtna hjälpmedel är en räknare, formelsamling i matematisk statistisk utgiven av MAI, och ett ytterligare formelblad (framsida och baksida).

Betygsgränser 3: minst 8 poäng, 4: minst 11.5 poäng, 5: minst 15 poäng (av 18)

(1) För de två händelserna A och B gäller:

$$P(A) = 0.4, \quad P(B) = 0.5\,.$$

Bestäm $P(A \cup B)$ och $P(A \cap B)$ då:

(a) A och B är oberoende händelser, (1p)

(b) A och B är disjunkta (oförenliga) händelser, (1p)

(c) sannolikheten för att A men inte B inträffar är 0.3, d.v.s. $P(A \cap B^c) = 0.3$. (1p)

(2) Antag att sannolikheten att en godtyckligt vald person i Sverige kommer att träffas av blixten under det närmaste året är 10^{-7}.

(a) Om det finns 9 miljoner invånare i Sverige, hur stor är då sannolikheten att minst tre personer kommer att träffas av blixten under det närmaste året? (1.5p)

(b) Om blixtnedslagen under de därpå följande tre åren är oberoende av dem under det närmaste året, hur stor är då sannolikheten att sammanlagt minst två personer kommer att träffas av blixten under de närmaste fyra åren? (1.5p)

Ledning: Om $X \sim Bin(n,p)$ så är X approximativt $Po(np)$-fördelad. Om $X_1 \sim Bin(n_1,p)$ och $X_2 \sim Bin(n_2,p)$ är oberoende, så är $X_1 + X_2 \sim Bin(n_1+n_2,p)$.

(3) Livslängden hos en viss typ av elektroniska komponent modelleras med hjälp av en slumpvariabel X med väntevärde $E[X] = 5$ dygn och varians σ^2. Livslängden hos olika komponenter antas vara oberoende.

(a) Beräkna sannolikheten att den sammanlagda livslängden hos 30 komponenter ligger mellan 120 och 180 dygn om $\sigma = 4.5$. Använd centrala gränsvärdessatsen. (1.5p)

(b) Man vill garantera att sannolikheten att den sammanlagda livslängden hos 30 komponenter ligger mellan 120 och 180 dygn är minst 0.9. Vad är det största värdet på σ som gör att detta uppfylls? (1.5p)

(4) I samband med en trafikomläggning med syfte att förkorta restiderna ville man undersöka förändringen i restider till arbete. Man bad därför 10 personer registrera restiderna (i minuter) dels en viss dag före omläggningen, dels en viss dag efter omläggingen. Resultatet blev:

Person nr	1	2	3	4	5	6	7	8	9	10
Före	20	26	23	24	48	57	27	18	14	28
Efter	18	23	18	25	44	51	26	22	11	30

Har omläggningen signifikant förkortat restiden? Testa på 5% -nivån under antagande att data är normalfördelat.

(a) Ange det generella testet (hypoteser, test variabel, kritiskt område, beslut) som ska användas. (1.5p)

(b) Genomför testet. (1.5p)

(5) Rent järn är relativt mjukt. Den betydligt hårdare legeringen stål erhålls genom att man tillsätter små mängder kol. För att skatta sambandet mellan kolhalt och hårdhet har man mätt hårdheten hos ett antal prover med i förväg valda kolhalter. (Hårdheten har angivits i diamond pyramid hardness, DPH. Kolhalten mäts i procent.) Resultatet framgår av nedanstående tabell.

Kolhalt i % (x_i)	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4
Hårdhet (y_i)	66	116	129	175	209	238	269	301

Man anpassar en modell till data: $Y_i = \alpha + \beta x_i + \varepsilon_i$, där $\varepsilon_i \sim N(0, \sigma^2)$ är sinsemellan oberoende

(a) Beräkna ett tvåsidigt 99% konfidensintervall för β. (1.5p)

(b) Beräkna ett 99% prediktionsintervall för hårdheten när kolhalten är 1.3 procent. (1.5p)

Räknehjälp: Vissa av följande kvadratsummor och medelvärden kan vara till nytta. Vi har betecknat kolhalten med x och hårdheten med y.

$$\bar{x} = 0.7, \quad \bar{y} = 187.875,$$

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 1.680, \quad \sum_{i=1}^{8} (y_i - \bar{y})^2 = 45988.9, \quad \sum_{i=1}^{8} (x_i - \bar{x})(y_i - \bar{y}) = 277.1.$$

(6) Man vill undersöka om utlåningsfrekvensen för ett bibliotek varierar med veckodag. Under en slumpmässigt vald vecka erhölls följande resultat.

veckodag	mån	tis	ons	tor	fre
# utlånade böcker	135	108	120	114	146

Testa på signifikansnivån $\alpha = 0.05$ huruvida utlåningen varierar med veckodag. (3p)

Lösningar

(1) (a)

$$P(A \cap B) = P(A) \cdot P(B) = 0.4 \cdot 0.5 = 0.2 \,.$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - 0.2 = 0.7 \,.$$

(b)

$$P(A \cap B) = 0, P(A \cup B) = P(A) + P(B) = 0.9 \,.$$

(c)

$$P(A \cap B) = P(A) - P(A \cap B^c) = 0.4 - 0.3 = 0.1 \,.$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - 0.1 = 0.8 \,.$$

(2) Låt X vara antalet som träffas av blixten ett typiskt år. Då är

$$X \sim Bin(9 \cdot 10^6, 10^{-7}) \approx Po(0.9) \,.$$

(a)

$$P(X \geq 3) = 1 - \sum_{k=0}^{2} P(X = k) \approx 1 - e^{-0.9}(1 + 0.9 + 0.9^2/2) = 0.0629 \,.$$

(b) Låt X_i vara antalet som träffas av blixten under år $i = 1, 2, 3, 4$ och låt $Y = \sum_{i=1}^{4} X_i$. Då gäller att

$$Y \sim Bin(4 \cdot 9 \cdot 10^6, 10^{-7}) \approx Po(3.6) \,.$$

Den sökta sannolikheten ges av

$$P(Y \geq 2) = 1 - P(Y = 0) - P(X = 1) \approx 1 - e^{-3.6}(1 + 3.6) = 0.8743 \,.$$

(3) Låt $X_1, X_2, \ldots, X_{30}$ beteckna komponenternas livslängder i dygn. Vi har då att $X_1, X_2, \ldots, X_{30}$ är oberoende och likafördelade med $E[X_i] = 5$ dygn och $Var(X_i) = \sigma^2$. Låt Y vara den sammanlagda livslängden för 30 komponenter, dvs. $Y = X_1 + \ldots + X_{30}$.

(a) Eftersom Y är en summa av 30 st oberoende likafördelade stokastiska variabler kan vi använda oss av centrala gränsvärdessatsen för att beräkna sannolikheten att $120 \leq Y \leq 180$ enligt

$$\begin{aligned} P(120 \leq Y \leq 180) &= P\left(\frac{120 - 30 \cdot 5}{\sigma\sqrt{30}} \leq \frac{Y - 30 \cdot 5}{\sigma\sqrt{30}} \leq \frac{180 - 30 \cdot 5}{\sigma\sqrt{30}}\right) \\ &\approx P\left(\frac{-\sqrt{30}}{\sigma} \leq Z \leq \frac{\sqrt{30}}{\sigma}\right) \,. \end{aligned}$$

Här har vi $Z \sim N(0, 1)$ vilket tillsammans med $\sigma = 4.5$ och symmetriargument ger att

$$P(120 \leq Y \leq 180) \approx 2P\left(Y \leq \frac{\sqrt{30}}{\sigma}\right) - 1 = 2\,\Phi(1.22) - 1 = 0.78 \,.$$

(b) För att bestämma σ så att $P(120 \leq Y \leq 180) \geq 0.9$ använder vi oss av att enligt centrala gränsvärdessatsen

$$0.9 \leq P(120 \leq Y \leq 180) \approx 2\,\Phi\left(\frac{\sqrt{30}}{\sigma}\right) - 1\,.$$

Vi får $\frac{\sqrt{30}}{\sigma} \geq 1.645$ och således $\sigma \leq 3.33$.

(4) (a) Data utgör stickprov i par. Vi ansätter en modell där restiden X_i för person i före trafikomläggningen är normalfördelad med väntevärde μ_i och restiden Y_i för person i efter omläggningen är normalfördelad med väntevärde $\mu_i + \Delta$. Skillnaden $Z_i = Y_i - X_i$ blir då normalfördelad med väntevärde Δ. Sökt är alltså test (3) i Föreläsning 8 för $H_1 : \Delta < 0 = \Delta_0$ som ni har skrivit ner som svar.

(b) Den observerade genomsnittliga skillnaden är $\bar{z} = -1.7$ minuter och vi skattar stickprovsvarianserna för skillnaderna med $s^2 = 10.23$. Under $H_0 : \Delta = 0$ gäller att

$$T(Z) := \frac{\bar{Z}}{S_Z/\sqrt{10}} \sim t(10-1).$$

Vi förkastar alltså H_0 till förmån för $H_1 : \Delta < 0$ på 5% -nivån om $t(z) < -t_{0.05}(9) = -1.83$. Vi har $t(z) = -1.7/(10.23/\sqrt{10}) = -1.68$ och förkastar alltså inte H_0 - förkortningen av restiden är inte signifikant på 5% -nivån.

(5) (a) Enligt formlarna är skattningarna

$$\beta^*_{obs} = \frac{S_{xy}}{S_{xx}} = \frac{277.1}{1.680} = 164.94,$$

och

$$(\sigma^*_{obs})^2 = s^2 = \frac{S_{yy} - S_{xy}^2/S_{xx}}{8-2} = \frac{45988.9 - 277.1^2/1.680}{6} = 47.32,$$

med $8-2 = 6$ frihetsgrader. Så $s = \sqrt{47.32} = 6.878$. Ett 99% konfidensintervall för β är därför

$$\beta^*_{obs} \pm t_{0.005}(6) \cdot \frac{s}{\sqrt{S_{xx}}} = 164.94 \pm 3.7074 \cdot \frac{6.878}{\sqrt{1.680}} = 164.94 \pm 19.68$$

eller $(145.3, 184.6)$.

(b) Vi har att $\alpha^*_{obs} = \bar{y} - \beta^*_{obs}\bar{x} = 187.875 - 164.94 \cdot 0.7 = 72.417$. Ett 99% prediktionsintervall för hårdheten när kolhalten är 1.3%. ges därför av

$$\alpha^*_{obs} + \beta^*_{obs} \cdot 1.3 \pm t_{0.005}(6) \cdot s \cdot \sqrt{1 + \frac{1}{8} + \frac{(1.3-\bar{x})^2}{S_{xx}}}$$

$$= 72.417 + 164.94 \cdot 1.3 \pm 3.7074 \cdot 6.878 \cdot \sqrt{1 + \frac{1}{8} + \frac{0.6^2}{1.680}}$$

$$= 286.84 \pm 29.51$$

eller $(257.3, 316.4)$.

(6) Vi använder χ^2-testet på en given fördelning med 5 möjliga utfall dvs. $r = 5$.

Nollhypotesen är

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = 0.2\,.$$

Vi har totalt $n = 623$ observationer, $x_1 = 135$, $x_2 = 108$, $x_3 = 120$, $x_4 = 114$, $x_5 = 146$. Dessutom konstaterar vi $np_j = 623 \cdot 0.2 = 124.6 \geq 5$, för $j = 1, 2, 3, 4, 5$.

Vi får

$$\begin{aligned}
t &= \sum_{j=1}^{r} \frac{(x_j - np_j)^2}{np_j} \\
&= \frac{(135 - 124.6)^2}{124.6} + \frac{(108 - 124.6)^2}{124.6} + \frac{(120 - 124.6)^2}{124.6} + \frac{(114 - 124.6)^2}{124.6} \\
&\quad + \frac{(146 - 124.6)^2}{124.6} \\
&= 7.8266\,.
\end{aligned}$$

Det gäller att $\chi^2_{\alpha}(r-1) = \chi^2_{0.05}(4) = 9.94 > 7.8266$, som betyder att vi inte kan förkasta H_0 på $\alpha = 5\%$.