LINKÖPINGS UNIVERSITET

Matematiska institutionen

26 mars 2021, kl. 08.00-12.00
Examinator: Jörg-Uwe Löbus

Alla hjälpmedel är tillåtna.

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

(1) Anta att den dagliga variationen av priset av en aktie är en stokastisk variabel med väntevärde 0 och varians σ^2. Låt Y_0 vara priset av aktien idag och låt Y_n representera priset av aktien n dagar senare. Då är

$$Y_n = Y_{n-1} + X_n\,, \quad n \geq 1,$$

där $X_1, X_2, \ldots$ är oberoende och likafördelade stokastiska variabler med väntevärde 0 och varians σ^2. Anta att priset av aktien idag, Y_0, är 100. Om $\sigma^2 = 1$, vad är den approximativa sannolikheten att priset är högre än 107 om 50 dagar? Använd centrala gränsvärdessatsen. (3p)

(2) Anta att 1% av populationen är snarkare. Vi tar en person från populationen på måfå och vill veta om han/hon är snarkare eller inte. Testet man kan använda ger ett rätt svar med sannolikhet 0.99 oavsett man är snarkare eller inte.

(a) För en person tagen på måfå, bestäm sannolikheten att testet säger personen är snarkare. (1.5p)

(b) Anta att testet säger att den valda personen är snarkare. Hur stor är då sannolikheten att han/hon faktiskt är snarkare? (1.5p)

(3) Man har länge använt kedjor i brännugnar i cementanläggningar för att minska värmeförbrukning. Man studerar om kedja har samma effekt när billigare material med hög svavel- och klorinnehåll används. Man använder två oberoende stickprov från

$$N(\mu_x, \sigma^2) \quad \text{resp.} \quad N(\mu_y, \sigma^2)$$

där

utan kedjor $n_x = 16$, $\bar{x} = 6150\ kJ/kg$, $s_x = 80\ kJ/kg$

med kedjor $n_y = 14$, $\bar{y} = 5250\ kJ/kg$, $s_y = 75\ kJ/kg$.

(a) Ange formeln för konfidensintervallet för differensen $\mu_x - \mu_y$. (1p)

(b) Är det bra att använda kedjor? Förklara varför/varför inte genom att hitta ett 99% konfidensintervall för $\mu_x - \mu_y$, dvs. differensen i värmeförbrukningen i brännugnar med och utan kedjor. (2p)

(4) Man gör undersökningar vid badplatser för att få reda på om halten av gift från giftalger överstiger vissa gränsvärden. Om man kan visa att den förväntade nivån överstiger 0.6 sätter man upp en varningsskylt som rekommenderar folk att ej bada där och om man kan visa att den förväntade nivån överstiger 0.8 utfärdar man badförbud. Man observerar nivåerna

$$1.18 \quad 0.99 \quad 0.83 \quad 0.71 \quad 1.27 \quad 0.49 \quad 1.58 \quad 1.05 \ .$$

Antag att nivåernas värden är oberoende och varierar enligt normalfördelning. Följande info kan vara användbar: Stickprovsmedelvärdet är $\bar{x} = 1.0125$ och stickprovsstandardavvikelsen är $s = 0.341$.

(a) Ska man sätta upp varningsskylten? Genomför lämplig undersökning på nivå $\alpha = 0.01$. (1.5p)

(b) Ska man utfärda badförbud? Genomför lämplig undersökning på nivå $\alpha = 0.01$. (1.5p)

(5) I skidskytte ska man i varje skjutning träffa med fem skott. Tänk att man följer en skidskytt under 250 skjutningar både på träning och tävling och registrerar antalet träffar i varje skjutning. Man arbetar enligt modellen att skidskytten träffar varje skott med sannolikheten p oberoende av andra skott. I en skjutning blir alltså X=antalet träffade skott binomialfördelad med parametrarna $n = 5$ och p.

(a) Härled Maximum-Likelihoodskattningen av p. Beräkna ett värde för p^*_{obs} från observationerna nedan. (1.5p)

(b) Vi har fått följande observationer från de 250 skjutningarna:

Träffar	0	1	2	3	4	5
Antal obs.	25	65	33	21	60	46

Undersök med ett test på nivån 0.05 om vårt antagande om binomialf ördelning är korrekt. Tips: Använd uttrycket för skattningen från uppgift (a) på den okända parametern p. Om du gör det, hur blir då frihetsgraderna? Visa värdena som ingår i din räkning. (1.5p)

(6) Vissa tillverkade byggnadselement har en längd som är normalfördelade med väntevärden $\mu = 1.2$, i cm, och variansen $\sigma^2 = 0.04$, i cm^2. Antag att vi har tre byggnadselement med oberoende längder X_1, X_2, X_3.

(a) Bestäm fördelningens typ och parametrarnas värden för den stokastiska variabeln $S = X_1 + X_2 + X_3$. (1p)

(b) Beräkna sannolikheten att summan $X_1 + X_2 + X_3$ är mindre än 3.5. (2p)

Lösningar

(1) $Y_0 = 100$, $\sigma^2 = 1$, $Y_{50} = Y_{49} + X_{50} = \ldots = Y_0 + \sum_{i=1}^{50} X_i$. Centrala gränsvärdessatsen ger att

$$\frac{\sum_{i=1}^{50} X_i - 50E(X_1)}{\sqrt{50Var(X_1)}} = \frac{\sum_{i=1}^{50} X_i}{\sqrt{50}}$$

är approximativt $N(0,1)$-fördelad. Då är

$$\begin{aligned} P\left(Y_{50} > 107\right) &= P\left(\sum_{i=1}^{50} X_i > 7\right) \\ &= P\left(\frac{\sum_{i=1}^{50} X_i}{\sqrt{50}} > \frac{7}{\sqrt{50}}\right) \\ &\approx P(Z > 0.99) = 1 - \Phi(0.99) = 0.16\,, \end{aligned}$$

där $Z \sim N(0,1)$.

(2) Låt

$$E = \{\text{en person tagen på måfå är snarkare}\}$$

och

$$F = \{\text{testet säger att en person tagen på måfå är snarkare}\}\,.$$

Då är

$$P(E) = 0.01\,, \quad P(F|E) = P(F^c|E^c) = 0.99$$

och

$$P(F|E^c) = 1 - P(F^c|E^c) = 1 - 0.99 = 0.01\,.$$

(a) Det medför att

$$P(F \cap E) = P(F|E) \cdot P(E) = 0.99 \cdot 0.01\,,$$

$$P(F \cap E^c) = P(F|E^c) \cdot P(E^c) = 0.01 \cdot 0.99$$

och

$$P(F) = P(F \cap E) + P(F \cap E^c) = 0.99 \cdot 0.01 + 0.01 \cdot 0.99 = 0.0198\,.$$

(b) Vi får att

$$P(E|F) = \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.01 \cdot 0.99} = \frac{1}{2}\,.$$

(3) (a)

$$I_{\mu_x - \mu_y} = \left(\bar{x} - \bar{y} \pm t(n_x + n_y - 2)_{\alpha/2} \cdot s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\right)$$

där

$$s_p^2 = \frac{(n_x - 1) \cdot s_x^2 + (n_y - 1) \cdot s_y^2}{n_x + n_y - 2}$$

och

$$s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2\,, \quad s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2\,.$$

(b)

$$X_i \sim N(\mu_x, \sigma^2) \quad \text{och} \quad Y_j \sim N(\mu_y, \sigma^2).$$

Sammanvägd variansskattning:

$$s_p^2 = \frac{15 s_1^2 + 13 s_2^2}{28} = 6040.179\,.$$

Kvantilen $t_{0.005}(28) = 2.763$ fås ur t-tabellen. Vi får intervallet

$$I_{\mu_x - \mu_y} = \left(6150 - 5250 - 2.763 \cdot \sqrt{6040.179} \cdot \sqrt{\frac{1}{16} + \frac{1}{14}},\right.$$
$$\left. 6150 - 5250 + 2.763 \cdot \sqrt{6040.179} \cdot \sqrt{\frac{1}{16} + \frac{1}{14}}\right)$$
$$= (821.4, 978.6)\,.$$

Eftersom 0 inte är i intervallet verkar det rimligt att använda kedjor.

(4) (a) Värdena $x_1, \ldots, x_8$ är observationer av $X_1, \ldots, X_8$ som är oberoende och $N(\mu, \sigma)$-fördelade. Vi får $\bar{x} = 1.0125$ och $s = 0.341$. Hypoteserna

$$H_0 : \mu = 0.6 \qquad H_1 : \mu > 0.6$$

prövas med hjälp av

$$t = \frac{\bar{x} - 0.6}{s/\sqrt{8}} = 3.42\,.$$

H_0 förkasras eftersom $t > t_{0.01}(7) = 2.998$. Varningsskylt ska sättas upp.

(b) Hypoteserna

$$H_0 : \mu = 0.8 \qquad H_1 : \mu > 0.8$$

prövas med hjälp av

$$t = \frac{\bar{x} - 0.8}{s/\sqrt{8}} = 1.76\,.$$

H_0 förkasras inte eftersom $t < t_{0.01}(7) = 2.998$. Verkar inte finnas någon anledning att införa badförbud.

(5) (a) Likelihood-funktionen blir

$$L(p) = \prod_{i=1}^{m} \left[\binom{5}{x_i} \cdot p^{x_i} \cdot (1-p)^{5-x_i}\right] = \prod_{i=1}^{m} \binom{5}{x_i} \cdot p^{\sum_{i=1}^m x_i} \cdot (1-p)^{\sum_{i=1}^m (5-x_i)}\,.$$

Detta ger att

$$\ln L(p) = \ln\left(\prod_{i=1}^{m} \binom{5}{x_i}\right) + \ln(p) \cdot \sum_{i=1}^{m} x_i + \ln(1-p) \cdot \sum_{i=1}^{m} (5 - x_i)\,.$$

Genom att derivera och sätta derivatan lika med noll får man att

$$p^*_{\text{obs}} = \frac{\sum_{i=1}^m x_i}{5m}$$

vilket är ett maximum eftersom

$$\lim_{p\to 0} \ln L(p) = -\infty \quad \text{and} \quad \lim_{p\to 1} \ln L(p) = -\infty\,.$$

Sätter man in värdena fås att $p^*_{\text{obs}} = 664/(5 \cdot 250) = 0.5312$.

(b) Ett χ^2-test görs för att testa om binomialfördelnings- antagandet är korrekt. I uppgiften anges att $X \sim \text{Bin}(5, p)$ och därmed finns 6 kategorier, för vilka vi behöver beräkna sannolikheter.

$$p_i = P(X = i) = \binom{5}{i} \cdot p^i \cdot (1-p)^{5-i}, \quad i \in \{0, \ldots, 5\}.$$

Genom att sätta in det skattade värdet på p från (a) fås att

$$(p_0^*, \ldots, p_5^*) = (0.022643,\ 0.12829,\ 0.29072,\ 0.32942,\ 0.18663,\ 0.042295).$$

Teststorheten är

$$t = \sum_{i=0}^{5} \frac{(N_i - mp_i^*)^2}{mp_i^*} \approx 290 > 9.488 = \chi^2_{0.05}(4)$$

H_0 förkastas. Antagandet om binomialfördelning är mycket troligen falskt.

(6) (a) $S := X_1 + X_2 + X_3$ är normalfördelad med $\mu_S = E[S] = 3.6$, $\sigma_S^2 = Var(S) = 0.12$.

(b)

$$P(S < 3.5) = P\left(\frac{S - \mu_S}{\sigma_S} < \frac{3.5 - 3.6}{\sqrt{0.12}}\right) = P\left(Z < -0.29\right) = 1 - \Phi(0.29) = 0.386.$$