Vzorová písemná část zkoušky ze ZSM, jarní semestr 2009

Úkol 1.: Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozložení Ex(λ). Určete konstantu c tak, aby statistika $T = c(X_1 + \ldots + X_n)$ byla nestranným odhadem parametrické funkce $\frac{1}{\lambda}$.

Řešení: $\forall \lambda > 0: \frac{1}{\lambda} = E(T) = E\left(c\sum_{i=1}^{n} X_i\right) = c\sum_{i=1}^{n} E(X_i) = c\sum_{i=1}^{n} \frac{1}{\lambda} = cn\frac{1}{\lambda} \Rightarrow c = \frac{1}{n}$

Úkol 2.: Nechť X_1, X_2 je náhodný výběr z normálního rozložení se střední hodnotou 100 a směrodatnou odchylkou 5. Jaká je pravděpodobnost, že výběrový úhrn (tj. součet $X_1 + X_2$) se bude realizovat hodnotou aspoň 221?

Řešení:

$$P(X_1 + X_2 \geq 221) = 1 - P(X_1 + X_2 \leq 221) + P(X_1 + X_2 = 221) = 1 - P\left(M \leq \frac{221}{2}\right) =$$

$$= 1 - P\left(\frac{M - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{110{,}5 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - P\left(U \leq \frac{110{,}5 - 100}{\frac{5}{\sqrt{2}}}\right) = 1 - \Phi\left(\frac{10{,}5\sqrt{2}}{5}\right) = 1 - \Phi(2{,}97) =$$

$$= 1 - 0{,}9985 = 0{,}0015;$$

Úkol 3.: Je dán náhodný výběr rozsahu 6800 z dvourozměrného diskrétního rozložení, přičemž veličina X nabývá tří variant a veličina Y nabývá čtyř variant. Testová statistika pro test nezávislosti veličin X a Y nabyla hodnoty 1073,5.

a) Lze na asymptotické hladině významnosti 0,05 zamítnout hypotézu o nezávislosti veličin X a Y?

b) Jaký je Cramérův koeficient a jak ho lze interpretovat?

Řešení:

ad a) $W = \langle \chi^2_{0,95}((3-1)(4-1)), \infty) = \langle \chi^2_{0,95}(6), \infty) = \langle 12{,}592, \infty)$. Protože testová statistika se realizuje v kritickém oboru, nulovou hypotézu zamítáme na asymptotické hladině významnosti 0,05.

ad b) $V = \sqrt{\frac{K}{n(m-1)}} = \sqrt{\frac{1073{,}5}{6800\cdot(3-1)}} = \sqrt{\frac{1073{,}5}{13600}} = 0{,}281$. Mezi veličinami X a Y existuje jenom slabá závislost.

Úkol 4.: Ze 40 naměřených údajů byl vypočten průměr 30 a koeficient variace 0,2. Dodatečně bylo zjištěno, že všechny údaje byly nadhodnoceny o 2 jednotky. Jaká je správná hodnota průměru a správná hodnota rozptylu?

Řešení:

$$cv_p = \frac{s_p}{m_p} \Rightarrow s_p = cv_p \cdot m_p = 0{,}2 \cdot 30 = 6$$

Správný průměr: $m_S = m_P - 2 = 30 - 2 = 28$

Správný rozptyl je stejný jako původní rozptyl, tedy $6^2 = 36$.

Úkol 5.: Nechť $X_1, \ldots, X_{400}$ je náhodný výběr z N(μ,0,01). Je známo, že výběrový průměr se realizoval hodnotou 0,01. Na hladině významnosti 0,05 testujte hypotézu H_0: $\mu = 0$ proti pravostranné alternativě H_1: $\mu > 0$ pomocí p-hodnoty.

Řešení:

Realizace testové statistiky: $t_0 = \frac{m - c}{\frac{\sigma}{\sqrt{n}}} = \frac{10{,}01 - 10}{\frac{0{,}1}{20}} = 2$,

p-hodnota = $P(T_0 \geq t_0) = 1 - \Phi(t_0) = 1 - \Phi(2) = 1 - 0{,}97725 = 0{,}02275$, nulovou hypotézu tedy zamítáme na hladině významnosti 0,05.

Úkol 6.: Jsou dány tři nezávislé náhodné výběry o rozsazích 4, 3, 4. V prvním výběru jsou dvě hodnoty menší než medián vypočtený ze všech 11 hodnot, ve druhém výběru jsou všechny hodnoty větší než medián a ve třetím výběru je jedna hodnota stejná jako medián a zbylé tři jsou menší než medián. Lze na asymptotické hladině významnosti 0,05 zamítnout hypotézu, že všechny tři výběry pocházejí z téhož rozložení?

Řešení:

Jedná se o mediánový test. Testová statistika má tvar $Q_M = 4\sum_{j=1}^{r} \frac{P_j^2}{n_j} - n$, kde P_j je počet hodnot v j-tém výběru, které jsou větší nebo rovny mediánu vypočtenému ze všech n hodnot. Platí-li H_0, má statistika Q_M asymptoticky rozložení $\chi^2(r-1)$. H_0 tedy zamítneme na asymptotické hladině významnosti α, když $Q_M \geq \chi_{1-\alpha}{}^2(r-1)$.

V našem případě r = 3, $n_1 = 4$, $n_2 = 3$, $n_3 = 4$, n = 11, $P_1 = 2$, $P_2 = 3$, $P_3 = 1$. Dostáváme tedy $Q_M = 4\left(\frac{2^2}{4} + \frac{3^2}{3} + \frac{1^2}{4}\right) - 11 = 4(1 + 3 + 0{,}25) - 11 = 6$, $\chi_{1-\alpha}{}^2(r-1) = \chi_{0{,}95}{}^2(2) = 5{,}991$. Protože testová statistika se realizuje v kritickém oboru, na asymptotické hladině významnosti 0,05 zamítáme hypotézu, že všechny tři výběry pocházejí z téhož rozložení.