Zkouškový test z PMS

na vypracování máte 90 min

13.1.2017

.

T1 Mám T triček, z nichž Z je zelených, M je modrých a zbytek je černý. Náhodně vyberu k triček, jaká je pravděpodobnost, že z nich bude právě z zelených a c černých? Řešte pouze pro nenulové a smysluplné hodnoty konstant (vyhněte se trivialitám typu $k > T$ nebo $z > Z$). Stanovte hodnotu této pravděpodobnosti pro hodnoty $T = 10, Z = 3, M = 4, k = 7, z = 2, c = 1$.

T2 Trolů je v populaci třikrát více než trpaslíků a lidí je dvakrát více než trolů. Testem inteligence projde průměrně 50% lidí, 70% trpaslíků a pouze 10% trolů. Test všichni vyplňují nezávisle. Náhodně vyberu z populace dvě postavy.

a) Jaká je pravděpodobnost, že je to trol a člověk?

b) Jaká je pravděpodobnost, že obě dvě prošly testem?

c) První postava testem prošla. Jaká je pravděpodobnost, že to byl člověk?

T3 Náhodný vektor (X, Y) má rovnoměrné rozdělení na kruhu v $\mathbb{R}^2$ o poloměru 3 se středem v počátku. Zobrazíme jej funkcí $h(x, y) = x^2 + y^2$. Popište rozdělení veličiny $Z = h(X, Y)$ a najděte její střední hodnotu.

T4 Metodou maximální věrohodnosti odhadněte parametr Poissonova rozdělení λ, pokud výsledky 20 měření veličiny $X \sim \mathrm{Po}(\lambda)$ shrnuje tabulka:

hodnota	1	2	4	5	6
četnost	1	3	7	6	3

T5 Otestujte, zda jsou dospělí synové v průměru vyšší než jejich otcové. K tomu jsme náhodně vybrali 8 dvojic otec-syn, jejichž výšky v centimetrech udává tabulka.

otec	174	182	169	177	175	180	172	190
syn	179	180	171	184	175	188	171	194

Jako hladinu významnosti volte $\alpha = 0.05$.

Řešení

T1 Počet způsobů, jak vybrat z trik z počtu Z zelených trik je $\binom{Z}{z}$, obdobně pro jiné kombinace. Celkově dostáváme

$$P_1 = \frac{\binom{Z}{z}\binom{T-M-Z}{c}\binom{M}{k-z-c}}{\binom{T}{k}}, \quad P_2 = \frac{\binom{3}{2}\binom{3}{1}\binom{4}{4}}{\binom{10}{7}} = \frac{3 \cdot 3 \cdot 1}{120} = 0.075.$$

T2 Označíme-li procentuální zastoupení trpaslíků v populaci p_1 (to jest rovno pravděpodobnosti, že když náhodně vyberu jedince z populace, bude to trpaslík), dostáváme rovnici

$$p_1 + 3p_1 + 2 \cdot 3p_1 = 1 \Rightarrow p_1 = 0.1.$$

a) Označme nyní jevy "první postava je trpaslík, resp. trol, resp. člověk"po řadě D, resp. T, resp. H. Jistě platí $p_1 = P(D)$. U druhé postavy jsou tyto pravděpodobnosti stejné a protože vybírám postavy nezávisle, platí

$$P(\text{vybral jsem trola a človeka}) = P(T)P(H) + P(H)P(T) = 2 \cdot 0.3 \cdot 0.6 = 36\%.$$

b) Označíme I jev, že první postava prošla testem. Dle věty o úplné pravděpodobnosti máme

$$P(I) = P(I|D)P(D) + P(I|T)P(T) + P(I|H)P(H) = 0.7 \cdot 0.1 + 0.1 \cdot 0.3 + 0.5 \cdot 0.6 = 40\%.$$

c) $P(H|I) = \frac{P(I|H)P(H)}{P(I)} = \frac{0.3}{0.4} = 75\%.$

T3 Náhodná veličina Z bude zřejmě nabývat hodnot v rozmezí $[0, 9]$, její distribuční funkci tedy zkonstruujeme následovně

$$F_Z(z) = P\left(X^2 + Y^2 \le z\right) = \begin{cases} 0\,, & z < 0\,, \\ P\left(\sqrt{X^2+Y^2} \le \sqrt{z}\right)\,, & z \in [0, 9)\,, \\ 1\,, & z \ge 9\,. \end{cases}$$

Veličina $\sqrt{X^2 + Y^2}$ udává vzálenost náhodně vybraného bodu od středu kruhu. Tedy pro $z \in [0.9]$ je pravděpodobnost $P\left(\sqrt{X^2+Y^2} \le \sqrt{z}\right)$ rovna poměru obsahů soustředných kruhů o poloměrech $\sqrt{z}$ a 3. To jest $\frac{\pi(\sqrt{z})^2}{\pi 3^2} = \frac{z}{9}$. Celkově dostáváme

$$F_Z(z) = \begin{cases} 0\,, & z < 0\,, \\ \frac{z}{9}\,, & z \in [0, 9)\,, \\ 1\,, & z \ge 9\,, \end{cases}$$

tedy veličina Z má rovnoměrné rozdělení na intervalu $[0, 9]$. Zřejmě $\mathbb{E}Z = \frac{9-0}{2} = 4.5$.

T4 Nechť $X_1, ..., X_{20} \sim \text{Po}(\lambda)$ a $x_1, ..., x_{20} \in \mathbb{R}$ jsou po řadě jejich realizace. Pak pro věrohodnostní funkci $L(\lambda)$ platí

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{20} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \\ \log L(\lambda) &= \sum_{i=1}^{20}\left(\log e^{-\lambda} + \log\left(\frac{\lambda^{x_i}}{x_i!}\right)\right) \\ &= \sum_{i=1}^{20}\left(-\lambda + x_i \log\lambda - \log(x_i!)\right) \\ \frac{\partial}{\partial\lambda}\log L(\lambda) &= \sum_{i=1}^{20}\left(-1 + \frac{x_i}{\lambda}\right) = -20 + \frac{1}{\lambda}\sum_{i=1}^{20} x_i. \end{aligned}$$

Z tabulky realizací vyčteme $\sum_{i=1}^{20} x_i = 83$ a protože hledáme maximum funkce L, položíme derivaci rovnou nule. Tedy

$$
\begin{aligned}
0 &= -20 + \frac{1}{\lambda}\sum_{i=1}^{20} x_i \\
\lambda &= \frac{83}{20}.
\end{aligned}
$$

T5 Předně víme, že výška otce a syna jsou závislé veličiny. Chceme-li porovnávat jejich střední hodnoty, použijeme párový test. Protože lze očekávat, že výška jedince (či rozdíl výšek dvou jedinců) má normální rozdělení, použijeme párový T-test. Doplňme nyní tabulku o rozdíly výšek otce a syna:

otec (Y_i)	174	182	169	177	175	180	172	190
syn (X_i)	179	180	171	184	175	188	171	194
$Z_i = X_i - Y_i$	5	-2	2	7	0	8	-1	4

Předpokládejme, že $Z_i \sim \mathrm{N}(\mu, \sigma^2)$. Testujeme hypotézu $H_0 : \mu \leq 0$ proti $\mu > 0$. Jde tedy o jednostranný test. Testovací statistika

$$
T = \frac{\overline{Z} - 0}{S_Z}\sqrt{n}
$$

má za platnosti nulové hypotézy Studentovo t_7-rozdělení. Dosazením hodnot Z_i nám vyjde $\overline{Z} = 2.875$, $S_Z = 3.72$ a $T = 2.186$. Jest $q_{t_7}(0.95) = 1.895 < T$, proto na hladině 0.05 zamítáme hypotézu o tom, že synové jsou průměrně stejně vysocí jako jejich otcové. Na hladině spolehlivosti 0.05 lze tvrdit, že průměrná výška synů je větší než průměrná výška jejich otců.