Zkouškový test z PMS

na vypracování máte 90 min

23.1.2017

.

T1 Ve skupině je 15 lidí, z toho 9 mužů a 6 žen. Jaká je pravděpodobnost, že v náhodně vybrané pětici budou/bude

a) všichni muži?

b) alespoň jeden muž?

c) více mužů než žen?

T2 Test na nemoc vyjde pozitivní ve 5% případů, pokud je člověk zdravý a negativní ve 10% případů, pokud je člověk nemocný. Pozitivní test má 20% populace.

a) Kolik je v populaci nemocných?

b) Mám pozitivní test. Jaká je pravděpodobnost, že jsem nemocný?

T3 Nechť X, Y jsou náhodné veličiny, pro které platí $X \sim \text{Alt}(0.3)$ $Y \sim \text{Alt}(0.5)$. Nechť je jejich korelační koeficient roven $\rho = 21^{-1/2}$. Určete $\mathbb{E}XY$ a $\mathbb{E}(X^2 + Y^2)$.

T4 Hodíme 50x čtyřstěnnou kostkou. Nalezněte pravděpodobnost, že součet hodů bude mezi 110 a 130.

T5 Chceme testovat, zda se určitý živočišný druh vyskytuje v daném regionu rovnoměrně. Region jsme proto rozdělili na oblasti (s různými rozlohami) a v nich naměřili četnosti daného druhu. Výsledky měření udává tabulka.

Oblast	A	B	C	D	E
Rozloha v km^2	20	50	30	30	15
četnost	56	103	87	76	51

Jaký je váš závěr? Jako hladinu významnosti volte $\alpha = 0.05$.

Řešení

T1 Počet všech možných kombinací, jak vybrat 5 lidí z 15 je $\binom{15}{5}$.

a) Pokud jsou v pětici jen muži, lze je vybrat $\binom{9}{5}$ způsoby. Tedy

$$P = \frac{\binom{9}{5}}{\binom{15}{5}}.$$

b) Jev, že v pětici je alespoň jeden muž je doplňkový k jevu, že v pětici jsou samé ženy. Tedy

$$P = 1 - \frac{\binom{6}{5}}{\binom{15}{5}}.$$

c) Příznivé stavy jsou pro nás, kdy je v pětici $3, 4$ nebo 5 mužů. Možností, jak vybrat do pětice 3 muže a dvě ženy je $\binom{9}{3}\binom{6}{2}$, obdobně pro 4 muže a 1 ženu. Máme tedy

$$P = \frac{\binom{9}{3}\binom{6}{2} + \binom{9}{4}\binom{6}{1} + \binom{9}{5}}{\binom{15}{5}}$$

T2 Označíme-li jevy, že náhodně vybraný člověk: je zdravý Z, je nemocný N, má pozitivní, resp. negativní test $+$, resp. $-$, potom lze ze zadání rovnou určit

$$\begin{aligned} P(+|Z) &= 0.05 \quad P(-|N) = 0.1 \\ P(-|Z) &= 0.95 \quad P(+|N) = 0.9 \\ P(+) &= 0.2 \quad P(Z) = 1 - P(N) \end{aligned}$$

a) Věta o úplné pravděpodobnosti říká

$$\begin{aligned} P(+) &= P(+|Z)P(Z) + P(+|N)P(N) \\ &= P(+|Z)(1 - P(N)) + P(+|N)P(N) \\ 0.2 &= 0.05 - 0.05P(N) + 0.9P(N) \\ P(N) &= \frac{3}{17} \end{aligned}$$

b) $P(N|+) = \frac{P(+|N)P(N)}{P(+)} = \frac{0.9 \cdot \frac{3}{17}}{0.2} = \frac{27}{34}$.

T3 Označíme-li $\operatorname{var} X = \sigma_X^2$ a $\operatorname{var} Y = \sigma_Y^2$, máme

$$\rho_{X,Y} = \frac{\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y}{\sigma_X \sigma_Y}$$

$$\mathbb{E}XY = \rho_{X,Y}\sigma_X\sigma_Y + \mathbb{E}X\mathbb{E}Y.$$

Platí $\mathbb{E}X = 0.3$, $EY = 0.5$, $\sigma_X^2 = 0.21$, $\sigma_Y^2 = 0.25$, tedy

$$\mathbb{E}XY = \frac{\sqrt{0.21 \cdot 0.25}}{\sqrt{21}} + 0.3 \cdot 0.5 = 0.2$$

Dále $\mathbb{E}X^2 = 0^2 \cdot 0.7 + 1^2 \cdot 0.3 = 0.3$, $\mathbb{E}Y^2 = 0^2 \cdot 0.5 + 1^2 \cdot 0.5 = 0.5$, tedy $\mathbb{E}(X^2 + Y^2) = 0.8$.

T4 Označme $X_1, ..., X_{50}$ náhodné veličiny určující počet bodů, které padly na jednotlivých kostkách. Veličiny jsou nezávislé a stejně rozdělené s parametry $\mathbb{E}X_i = \mu = \frac{1}{4}(1+2+3+4) = 2.5$ a $\operatorname{var} X_i = \sigma_X^2 = EX^2 - (\mathbb{E}X)^2 = \frac{1}{4}(1^2+2^2+3^3+4^2) - 6.25 = 1.25$. Podle centrální limitní věty tedy bude mít veličina

$$Z = \frac{\sum_{i=1}^{50} X_i - 50 \cdot \mu}{\sqrt{50\sigma_X^2}} = \frac{\sum_{i=1}^{50} X_i - 125}{\sqrt{62.5}}$$

téměř normované normální rozdělení. Počítáme tedy

$$P\left(110 < \sum_{i=1}^{50} X_i < 130\right) = P\left(\frac{110-125}{\sqrt{62.5}} < Z < \frac{130-125}{\sqrt{62.5}}\right) = P\left(Z < \frac{\sqrt{10}}{5}\right) - P\left(Z < \frac{-3\sqrt{10}}{5}\right) =$$

$$\Phi\left(\frac{\sqrt{10}}{5}\right) - \left(1 - \Phi\left(\frac{3\sqrt{10}}{5}\right)\right) = 0.707.$$

T5 Jde o test parametru multinomického rozdělení. Parametry p_i', $i \in \{1,2,3,4,5\}$ odpovídají poměru rozlohy dané oblasti ku rozloze regionu. Zároveň celkový počet zaznamenaných druhů v regionu je roven $n = 373$. Tabulku tedy můžeme doplnit následovně.

oblast	A	B	C	D	E	celkem
rozloha v km^2	20	50	30	30	15	145
X_i (četnost)	56	103	87	76	51	373
p_i'	0.138	0.345	0.207	0.207	0.103	1
np_i (očekávaná četnost)	51.45	128.62	77.17	77.17	38.59	373
$\frac{(X_i-np_i)^2}{np_i}$ (přírůstek k χ^2)	0.402	5.103	1.252	0.018	3.991	10.766

Vektor $(X_1, ..., X_5)$ má tedy rozdělení $\mathrm{Mul}(n, p_1, ..., p_5)$ a nulová hypotéza zní

$$H_0 : \forall i \in \{1, ..., 5\} : \; p_i = p_i'.$$

Za platnosti H_0 má testovací statistika

$$T = \sum_{i=1}^{5} \frac{(X_i - np_i)^2}{np_i}$$

rozdělení χ_4^2. Z tabulek vyčteme $q_{\chi_4^2}(0.95) = 9.488 < T = 10.77$, proto na hladině 0.05 zamítáme hypotézu o tom, že se daný živočišný druh vyskytuje v regionu rovnoměrně.