Řešení vzorové písemné části zkoušky z Výpočetní statistiky

Úkol 1.: (8 bodů) Mezi atlety je rozšířen názor, že o vítězství rozhoduje přidělení běžecké dráhy. Proto bylo sledováno 160 závodů nejvyšší úrovně na světě (stadiony s 8 drahami). Výsledky jsou uvedeny v tabulce, která udává počet vítězství na jednotlivých drahách:

Číslo dráhy	1	2	3	4	5	6	7	8
Počet vítězství	23	21	17	22	19	18	24	16

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že přidělení běžecké dráhy nemá vliv na vítězství.

Řešení: Kdyby nezáleželo na dráze, byla by pravděpodobnost vítězství pro každou dráhu rovna 1/8. Výpočty uspořádáme do tabulky:

j	n_j	p_j	np_j	$(n_j - np_j)^2$	$(n_j - np_j)^2/ np_j$
1	23	1/8	20	9	9/20
2	21	1/8	20	1	1/20
3	17	1/8	20	9	9/20
4	22	1/8	20	4	4/20
5	19	1/8	20	1	1/20
6	18	1/8	20	4	4/20
7	24	1/8	20	16	16/20
8	16	1/8	20	16	16/20

$K = 3$, $r = 8$, $p = 0$, $\chi^2_{0,95}(7) = 14{,}067$. Protože $K < 14{,}067$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Úkol 2.: (8 bodů) Z realizace náhodného výběru rozsahu 12 jsme zjistili, že výběrový průměr nabyl hodnoty 3,8 a výběrový rozptyl 0,4. Dodatečně bylo zjištěno, že všechny údaje byly podhodnoceny o 1 jednotku. O kolik procent se změnil výběrový koeficient variace?

Řešení:

Původní výběrový koeficient variace: $cv_P = \dfrac{s_P}{m_P} = \dfrac{0{,}6325}{3{,}8} = 0{,}1664$

Správný průměr: $m_S = m_P + 1 = 3{,}8 + 1 = 4{,}8$

Správný rozptyl je stejný jako původní rozptyl, tedy 0,4.

Správný výběrový koeficient variace: $cv_S = \dfrac{s_S}{m_S} = \dfrac{0{,}6325}{4{,}8} = 0{,}1318$

Výběrový koeficient variace se změnil o $\left(\dfrac{0{,}1318 \cdot 100}{0{,}1664} - 100\right)\% = -20{,}8\%$

Úkol 3.: (8 bodů) Pomocí K-W testu testujeme hypotézu, že tři nezávislé náhodné výběry o rozsazích 4, 5, 5 pocházejí z téhož rozložení. Součet pořadí hodnot v 1. výběru je 25 a ve 2. výběru 39. Lze nulovou hypotézu na asymptotické hladině významnosti 0,05 zamítnout?

Řešení: $n = 14$, $T_1 + T_2 + T_3 = n(n+1)/2 = 14.15/2 = 105$, $T_3 = 105 - 25 - 39 = 41$

$$Q = \frac{12}{n(n+1)} \sum_{j=1}^{r} \frac{T_j^2}{n_j} - 3(n+1) = \frac{12}{14 \cdot 15}\left(\frac{25^2}{4} + \frac{39^2}{5} + \frac{41^2}{5}\right) - 3 \cdot 15 = 0{,}5229$$

Kritický obor pro K-W test má tvar $W = \langle \chi^2_{1-\alpha}(r-1); \infty) = \langle \chi^2_{0,95}(2); \infty) = \langle 5{,}991; \infty)$.

Testová statistika se nerealizuje ve W, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Úkol 4.: (8 bodů) Nechť $X_1, ..., X_{16}$ je náhodný výběr z N(-2, 9). Jaká je pravděpodobnost, že výběrový průměr nabude hodnoty aspoň -2,3?

Řešení:

$$P(M \geq -2{,}3) = 1 - P\left(\frac{M+2}{\frac{3}{4}} \leq \frac{-2{,}3+2}{\frac{3}{4}}\right) = 1 - P(U \leq -0{,}4) = 1 - \Phi(-0{,}4) = \Phi(0{,}4) = 0{,}65542$$

Hledaná pravděpodobnost je asi 65,5%.

Úkol 5.: (8 bodů) Z realizace náhodného výběru rozsahu 9, který pochází z rozložení $N(\mu, \sigma^2)$, byl vypočten výběrový průměr m = 15 a výběrový rozptyl $s^2 = 36$. Najděte 95% empirický interval spolehlivosti pro neznámou střední hodnotu μ.

Řešení: $d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 15 - \frac{6}{\sqrt{9}} t_{0,975}(8) = 15 - \frac{6}{3} 2{,}306 = 10{,}388$

h = 19,612, tedy $10{,}388 < \mu < 19{,}612$ s pravděpodobností aspoň 0,95.