Řešení písemné zkoušky z předmětu Aplikovaná statistika I, 27.1.2016 (pro antropology)

Příklad 1.: V následující tabulce jsou uvedeny počty správně vyřešených příkladů u přijímací zkoušky z matematiky a jejich absolutní četnosti.

x_j	0	1	2	3	4
n_j	5	10	16	18	13

a) Vypočtěte průměr a medián počtu správně vyřešených příkladů.
b) Vypočtěte směrodatnou odchylku počtu správně vyřešených příkladů.
Řešení:
ad a)
Průměrný počet správně vyřešených příkladů: :

$$m = \frac{1}{n}\sum_{j=1}^{r} n_j x_{[j]} = \frac{1}{62}(0.5 + 1.10 + 2.16 + 3.48 + 4.13) = \frac{148}{62} = 2{,}39$$

Medián počtu správně vyřešených příkladů: $x_{0,50} = \frac{x_{(31)} + x_{(32)}}{2} = \frac{2+3}{2} = 2{,}5$

ad b) Rozptyl počtu správně vyřešených příkladů:

$$s^2 = \frac{1}{n}\sum_{j=1}^{r} n_j (x_{[j]} - m)^2 =$$

$$= \frac{1}{62}\left[5\left(0 - \frac{74}{31}\right)^2 + 10\left(1 - \frac{74}{31}\right)^2 + 16\left(2 - \frac{74}{31}\right)^2 + 18\left(3 - \frac{74}{31}\right)^2 + 13\left(4 - \frac{74}{31}\right)^2\right] = 1{,}46$$

Směrodatná odchylka počtu správně vyřešených příkladů: $s = \sqrt{1{,}46} = 1{,}21$

Příklad 2.: U 10 dětí byla zjištěna výška v cm (znak X) a hmotnost v kg (znak Y): (132, 25), (126, 23), (138, 32), (123, 18), (135, 26), (132, 23), (124, 33), (138, 33), (135, 28), (128, 26).
Pro úsporu času máte uvedeny tyto součty:

$$\sum_{i=1}^{10} x_i = 1311, \sum_{i=1}^{10} y_i = 267, \sum_{i=1}^{10} x_i^2 = 172151, \sum_{i=1}^{10} y_i^2 = 7345, \sum_{i=1}^{10} x_i y_i = 35128.$$

Vypočtěte a interpretujte koeficient korelace výšky a hmotnosti dětí.
Řešení:

$$r_{12} = \frac{s_{12}}{s_1 s_2} = \frac{\frac{1}{10}\cdot 35128 - \frac{1}{10}\cdot 1311 \cdot \frac{1}{10}\cdot 267}{\frac{1}{10}\cdot 172151 - \left(\frac{1}{10}\cdot 1311\right)^2} = 0{,}5063$$

Mezi výškou a hmotností existuje středně silná přímá lineární závislost – čím větší výška, tím větší hmotnost.

Příklad 3.: Určitý elektronický přístroj má dobu životnosti, která se řídí exponenciálním rozložením se střední hodnotou 5000 hodin. Jaká je pravděpodobnost, že doba životnosti tohoto přístroje bude více než 1000 hodin?
Řešení: $X \sim Ex(0{,}0002)$.

$$P(X > 1000) = 1 - P(X \le 1000) = e^{-0{,}0002\cdot 1000} = e^{-0{,}2} = 0{,}8187$$

Příklad 4.: Hmotnost vajec je náhodná veličina, která se řídí normálním rozložením se střední hodnotou $\mu = 50$ g a směrodatnou odchylkou $\sigma = 5$ g.
Jaká je pravděpodobnost, že náhodně vybrané vejce bude mít hmotnost nejvýše 53 g?
Řešení: $X \sim N(50, 25)$

$$P(X \le 53) = P\left(\frac{X-50}{5} \le \frac{53-50}{5}\right) = P(U \le 0{,}6) = \Phi(0{,}6) = 0{,}72575$$

Příklad 5.: Náhodná veličina X má pravděpodobnostní funkci
$\pi(-2) = 0{,}1$; $\pi(-1) = 0{,}2$; $\pi(0) = 0{,}2$; $\pi(1) = 0{,}4$; $\pi(2) = 0{,}1$. Vypočtěte:
a) pravděpodobnost toho, že veličina X nabude hodnot z intervalu $\langle -1;1 \rangle$
b) střední hodnotu a rozptyl náhodné veličiny X.
Řešení:
ad a) $P(-1 \le X \le 1) = \pi(-1) + \pi(0) + \pi(1) = 0{,}2 + 0{,}2 + 0{,}4 = 0{,}8$
ad b) $E(X) = -2 \cdot 0{,}1 + (-1) \cdot 0{,}2 + 0 \cdot 0{,}2 + 1 \cdot 0{,}4 + 2 \cdot 0{,}1 = 0{,}2$
$D(X) = (-2)^2 \cdot 0{,}1 + (-1)^2 \cdot 0{,}2 + 0^2 \cdot 0{,}2 + 1^2 \cdot 0{,}4 + 2^2 \cdot 0{,}1 - 0{,}2^2 = 1{,}36$

Příklad 6.: Máme čtyři nezávislé náhodné výběry postupně z rozložení $N(\mu_1,\sigma^2)$, $N(\mu_2,\sigma^2)$, $N(\mu_3,\sigma^2)$, $N(\mu_4,\sigma^2)$, přičemž každý z nich má rozsah 6. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot, je-li známo, že skupinový součet čtverců je 46,5 a reziduální 67,5.
Řešení: $n = 24$, $r = 4$, $S_A = 46{,}5$, $S_E = 67{,}5$, $f_A = r - 1 = 3$, $f_E = n - r = 20$,
$$F_A = \frac{S_A/f_A}{S_E/f_E} = \frac{46{,}5/3}{67{,}5/20} = \frac{15{,}5}{3{,}375} = 4{,}593.$$ Protože testová statistika je větší než kvantil
$F_{0,95}(3,20) = 3{,}0984$, nulovou hypotézu zamítáme na hladině významnosti 0,05.

Příklad 7.: Jestliže výrobní linka na výrobu kuličkových ložisek pracuje správně, pak hmotnost kuliček je normálně rozložená náhodná veličina se střední hodnotou 4 g a směrodatnou odchylkou 0,1 g. Po seřízení linky bylo náhodně vybráno 25 kuliček. Jejich průměrná hmotnost činila 4,038 g. Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota hmotnosti kuliček se nezměnila proti oboustranné alternativě.
Řešení:
$X_1, \ldots, X_{25} \sim N(\mu, 0{,}1^2)$, H_0: $\mu = 4$ proti H_1: $\mu \neq 4$, $\alpha = 0{,}05$
a) Testování pomocí intervalu spolehlivosti:

$$d = m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} = 4{,}038 - \frac{0{,}1}{5} u_{0{,}975} = 4{,}038 - \frac{0{,}1}{5} 1{,}96 = 3{,}9998$$

$$h = m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} = 4{,}038 + \frac{0{,}1}{5} u_{0{,}975} = 4{,}038 + \frac{0{,}1}{5} 1{,}96 = 4{,}0772$$

Protože číslo $c = 4$ leží v intervalu (3,9998; 4,0772), H_0 nezamítáme na hladině významnosti 0,05.
b) Testování pomocí kritického oboru:
Vypočteme realizaci testové statistiky:

$$t_0 = \frac{m-c}{\frac{\sigma}{\sqrt{n}}} = \frac{4{,}038-4}{\frac{0{,}1}{5}} = 1{,}9$$

a stanovíme kritický obor: $W = \left(-\infty, -u_{0.975}\right\rangle \cup \left\langle u_{0.975}, \infty\right) = \left(-\infty, -1{,}96\right\rangle \cup \left\langle 1{,}96, \infty\right)$

Protože testová statistika se nerealizuje v kritickém oboru, na hladině významnosti 0,05 nezamítáme H_0.

c) Testování pomocí p-hodnoty:

$$p = 2\min\{\Phi(t_0), 1-\Phi(t_0)\} = 2\min\{\Phi(1,9), 1-\Phi(1,9)\} = 2\min\{0,97128, 1-0,97128\} = 0,05744 .$$

Protože p-hodnota je větší než hladina významnosti 0,05, H_0 nezamítáme na hladině významnosti 0,05.

Příklad 8.: Měřením délky deseti válečků byly získány tyto hodnoty (v mm): 5,38 5,36 5,35 5,40 5,41 5,34 5,29 5,43 5,42 5,32. Předpokládáme, že uvedené hodnoty jsou číselné realizace náhodného výběru rozsahu 10 z rozložení $N(\mu, \sigma^2)$, kde parametry μ a σ^2 neznáme. Sestrojte 95% empirický interval spolehlivosti pro směrodatnou odchylku σ.

Řešení:

$$d = \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}} = \sqrt{\frac{9 \cdot 0,00211}{\chi^2_{0,975}(9)}} = \sqrt{\frac{9 \cdot 0,00211}{19,023}} = 0,0316mm$$

$$h = \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}} = \sqrt{\frac{9 \cdot 0,00211}{\chi^2_{0,025}(9)}} = \sqrt{\frac{9 \cdot 0,00211}{2,7}} = 0,0839mm$$

Znamená to, že 0,0316 mm < σ < 0,0839 mm s pravděpodobností aspoň 0,95.