Cvičení 11: Nerovnosti, Popisná statistika a normální rozdělení

Teorie:

Datový soubor tvoří naměřené hodnoty: $x_1, x_2, \dots, x_n$. Po seřazení označujeme

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}.$$

Aritmetický průměr: $\bar{x} = m = \frac{1}{n}\sum_{i=1}^{n} x_i$.
Modus $\hat{x}$ je nejčastější hodnota znaku. Průměrná odchylka: $o = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$.
Rozptyl $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, směrodatná odchylka $s = \sqrt{s^2}$.
Modifikovaný rozptyl $\frac{n}{n-1}s^2$.
Centrované hodnoty $x_i - \bar{x}$, , standardizované hodnoty $\frac{x_i - \bar{x}}{s}$.
p-**kvantil** $(0 < p < 1)$ je

$$\tilde{x}_p = \begin{cases} x_{([np]+1} & \text{pro } np \notin \mathbb{Z} \\ x_{(np)} + x_{(np)+1} & \text{pro } np \in \mathbb{Z}, \end{cases}$$

kde $[a]$ značí celou část čísla a. Speciálně $\tilde{x}_{0,5}$ je medián, $\tilde{x}_{0,25}$, resp. $\tilde{x}_{0,75}$, dolní, resp. horní kvartil, $\tilde{x}_{0,1}, \tilde{x}_{0,2}, \dots, \tilde{x}_{0,9}$ decily apod.

Mezikvartilové rozpětí je definováno jako $Q = \tilde{x}_{0,75} - \tilde{x}_{0,25}$.

Pro **dvourozměrný datový soubor** $[x_i, y_i]$, kde $1 \leq i \leq n$ definujeme *kovarianci prvního a druhého znaku* jako

$$s_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

a koeficient korelace mezi prvním a druhým znakem jako $r_{12} = \frac{s_{12}}{s_x s_y}$.

Krabicový diagram (box plot): dolní a horní strana základního obdélníka (krabice) odpovídá dolnímu a hornímu kvartilu, vodorovná čára uvnitř krabice mediánu (výška krabice je tedy mezikvartilové rozpětí). Dolní svislá úsečka (dolní fous) odpovídá hodnotám, které leží „pod krabicí“ ve vzdálenosti nejvýše 1,5 násobku její výšky, obdobně horní fous. Mimo fousy se znázorňují ostatní body (tzv. odlehlá pozorování). Křížek uvnitř krabice znázorňuje aritmetický průměr.

Příklad 152. Vypočtěte průměr a rozptyl

a) centrovaných hodnot

b) standardizovaných hodnot.

Příklad 153. Dokažte, že pro rozptyl platí vztah $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$ a odvoďte analogický vztah pro modifikovaný rozptyl.

Příklad 154. Dokažte

a) pro $n = 2$,

b) obecně,

že pro koeficient korelace platí tzv. Cauchyova nerovnost $-1 \leq r_{12} \leq 1$.

Příklad 155. Byly naměřeny následující hodnoty nějakého znaku

$$10; 7; 7; 8; 8; 9; 10; 9; 4; 9; 10; 9; 11; 9; 7; 8; 3; 9; 8; 7.$$

Určete aritmetický průměr, medián, kvartily, rozptyl a příslušný krabicový diagram.

Příklad 156. Mějme nezápornou náhodnou veličinu X se střední hodnotou μ.

1. Bez dalších informací o rozdělení X odhadněte $P(X > 3\mu)$.
2. Víte-li, že $X \sim \mathrm{Ex}(\frac{1}{\mu})$, vypočtěte $P(X > 3\mu)$.

Příklad 157. Ke každému jogurtu běžné značky je náhodně (rovnoměrně) přibalen obrázek některého z 26 hokejových mistrů světa. Kolik jogurtů si fanynka Věrka musí koupit, aby s pravděpodobností 0,95 získala alespoň 5 kartiček Jaromíra Jágra?

Příklad 158. Určete pravděpodobnost, že při 1200 hodech kostkou padne šestka alespoň 150 krát a nejvýše 250 krát

1. pomocí Čebyševovy nerovnosti,
2. pomocí de Moivre-Laplaceovy věty.

Příklad 159. Průměrná rychlost větru je na určitém místě 20 km/hod.

- Bez ohledu na rozdělení rychlosti větru jako náhodné veličiny určete pravděpodobnost, že při jednom pozorování rychlost větru nepřesáhne 60 km/h.
- Určete interval, v němž se bude rychlost větru nacházet s pravděpodobností alespoň 0,9, víte-li navíc, že směrodatná odchylka $\sigma = 1$ km/hod.

Příklad 160. Na FI je 10% studentů s prospěchem do 1,2. Jak velkou skupinu je třeba vybrat, aby s pravděpodobností 0,95 v ní bylo 8-12% studentů s prospěchem do 1,2? Úlohu řešte zvlášť pomocí Čebyševovy a zvlášť pomocí Moivre-Laplaceovy věty.

Příklad 161. Dokažte, že pro kvantily normovaného normálního rozdělení platí vztah

$$-u_{\frac{\alpha}{2}} = u_{1-\frac{\alpha}{2}}.$$