Popisná statistika

základní soubor X výběrový soubor

Naměřili jsme n hodnot

$$x_1, x_2, \dots, x_n,$$

počet prvků souboru je tzv. **rozsah** souboru. Pro lepší zpracování data uspořádáme:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

a dostaneme **uspořádaný soubor hodnot**

Míry polohy

Průměr (resp. výběrový, aritmetický průměr)

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

p-kvantil (výběrový p-kvantil)

$$\tilde{x}_p = \begin{cases} x_{([np]+1)} & np \neq [np] \\ \frac{1}{2}(x_{(np)} + x_{(np+1)}) & np = [np] \end{cases},$$

kde $[a]$ značí celou část z a a $0 < p < 1$.

Míry variability

Rozptyl (výběrový rozptyl)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \right)$$

Kvartilové rozpětí

$$r_Q = \tilde{x}_{0,75} - \tilde{x}_{0,25}$$

Krabicový diagram (box plot, box and whisker plot, vousatá krabička)

„Krabička“ je ohraničena hodnotami kvartilů a je zobrazen medián. „Vousky“ znázorňují hodnoty, které nejsou od jednotlivých kvartilů vzdálené o více jak 1,5 násobek R_Q. Jednotlivě jsou vyznačena pozorování, která jsou ve větší vzdálenosti.

1. Byly naměřeny hodnoty nějakého jevu:

   10; 7; 7; 8; 8; 9; 10; 9; 4; 9; 10; 9; 11; 9; 7; 8; 3; 9; 8; 7
   Určete průměr, medián, kvartily, rozptyl, mezikvartilové rozpětí, hodnoty znázorněte pomocí krabicového diagramu a zakreslete výběrovou distribuční funkci.

Náhodný výběr

Náhodným výběrem (rozsahu n) nazýváme posloupnost n stochasticky nezávislých náhodných veličin $X_1, X_2 \dots, X_n$, které mají stejné rozložení, tedy $X_i \sim F(x_i)$, $i = 1, 2, \dots, n$.

Pozn.: Prakticky se s náhodným výběrem setkáváme při nezávislém vícenásobném opakování téhož pokusu.

Statistika: Náhodná veličina, která vznikne transformací náhodného výběru, se nazývá statistika.

Významné statistiky:

- *Výběrový průměr*
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- *Výběrový rozptyl*
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$
- *Výběrová směrodatná odchylka*
$$S = \sqrt{S^2}$$
- *Výběrová kovariance*
$$S_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2)$$

2. Odvoďte pravděpodobnostní funkci náhodného výběru z alternativního rozložení $A(\theta)$.
[$\pi(\boldsymbol{x}) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$ pro $x_i = 0, 1$, $i = 1, \dots n$ a $\pi(\boldsymbol{x}) = 0$ jinak]

3. Odvoďte hustotu náhodného výběru z normálního rozložení $N(\mu, \sigma^2)$. [$N_n(\mu\mathbf{1}, \sigma^2\mathbb{I})$]

4. Nechť $X_1, X_2 \dots, X_n$ je náhodý výběr z rozložení, které má střední hodnotu μ a rozptyl σ^2. Vypočítejte střední hodnotu a rozptyl výběrového průměru $\overline{X}$ a střední hodnotu výběrového rozptylu S^2. [$E(\overline{X}) = \mu$, $D(\overline{X}) = \frac{\sigma^2}{n}$, $E(S^2) = \sigma^2$]

5. Odvoďte rozložení výběrového průměru $\overline{X}$, jestliže náhodný výběr pochází
   (a) z normálního rozložení $N(\mu, \sigma^2)$, [$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$]
   (b) z alternativního rozložení $A(\theta)$. [$\pi_*(\overline{x}) = \binom{n}{n\overline{x}}\theta^{n\overline{x}}(1-\theta)^{n-n\overline{x}}$ pro $\overline{x} = 0, 1/n, 2/n, \dots, 1$ a $\pi_*(\overline{x}) = 0$ jinak]

6. Nechť $(X_{11}, X_{21})', \dots, (X_{1n}, X_{2n})'$ je náhodný výběr z rozložení s vektorem středních hodnot $(\mu_1, \mu_2)'$ a kovariancí $C(X_{1i}, X_{2i}) = \kappa$, $i = 1, \dots, n$. Vypočťete střední hodnotu výběrové kovariance S_{12}. [$S_{12} = \kappa$]

7. Předpokládejme, že velký ročník na vysoké škole má výsledky ze statistiky normálně rozloženy kolem střední hodnoty 72 bodů se směrodatnou odchylkou 9 bodů. Určete pravděpodobnost, že
   (a) náhodně vybraný student bude mít výsledek nad 80 bodů [0,18673]
   (b) průměr výsledků náhodného výběru 10 studentů bude větší než 80 bodů. [0,00248]

8. Nechť $X_1, X_2 \ldots, X_n$ je náhodný výběr z rozložení $N(10, 4)$. Jaká je pravděpodobnost, že výběrový průměr nabude hodnoty nejvýše 9.5? Sledujte vliv rozsahu výběru na tuto pravděpodobnost.

$[\Phi(-0,25\sqrt{n})]$