**Poznámky k předmětu Aplikovaná statistika, 3.téma**

# 3. Charakteristiky náhodých veličin

Při praktické práci s daty často na začátku nemůžeme tušit, z jakého rozdělení (resp. jaké jsou parametry daného rozdělení) naše data pocházejí. Je proto dobré si nejprve udělat o datech určitou představu podle některých typických číselných údajů. Na základě dat samozřejmě získáme jen orientační představu o tom, jakých hodnot číselné charakteristiky nabývají. V této kapitole se budeme věnovat teoretickým číselným charakteristikám svázaným s rozdělením náhodné veličiny, nikoliv jejich odhadům!

- **Střední hodnota náhodné veličiny** (značíme $\mathbb{E}X, \mu$)
  patří mezi tzv. *míry polohy*. Udává „očekávanou" hodnotu, okolo které by se měla veličina pohybovat.
  Pro **diskrétní** náhodnou veličinu s pravděpodobnostní funkcí $p(\cdot)$ definujeme střední hodnotu následovně:
  $$\mathbb{E}X = \sum_i x_i \cdot p(x_i).$$
  Pro **spojitou** náhodnou veličinu s hustotou $f$ potom vzorcem
  $$\mathbb{E}X = \int_{\mathbb{R}} x \cdot f(x)\mathrm{d}x.$$
  VLASTNOSTI STŘEDNÍ HODNOTY
  1. $\mathbb{E}(aX + b) = a\,\mathbb{E}X + b,\ a, b \in \mathbb{R}$,
  2. pro náhodné veličiny $X, Y$ platí: $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$,
  3. pokud $\mathbb{P}(X \geq 0) = 1$, pak $\mathbb{E}X \geq 0$.

**Poznámka 1.** *Právě s pojmem střední hodnoty souvisí ony mediální informace typu „průměrný Čech vypije tolik a tolik piva" (za rok 2015 to bylo 146,6 l na osobu), „průměrný plat v České republice je…" (ve třetím čtvrtletí 2016 to bylo 27 220 Kč) a tak dále. Čtenář sám si jistě ve škole počítal, jaká známka mu vychází na vysvědčení, a pak se občas divil, že tomu výsledek od pedagoga neodpovídal. Je třeba si uvědomit, že samotná informace o tom, kolem které hodnoty se nám něco pohybuje, není dostatečně vypovídající. Jeden příklad ze života za všechny. Kamarádka hraje stolní tenis v mužské lize (protože je tak dobrá). A po jednom kole nám vyprávěla, že jsme se minule měly přijít podívat na soupeře – věkový průměr třicet (v době vzniku tohoto textu pro autorky věk velmi aktuální), čili lákavá podívaná. Ovšem pouze do okamžiku, než z ní vypadlo, že jednomu soupeři bylo sedmdesát a dvěma dalším deset. Něco takového prostě neumíme detekovat pomocí měr polohy. Je třeba přidat ještě alespoň jednu základní* míru variability*, rozptyl.*

- **Rozptyl náhodné veličiny** (značíme var$X, \sigma^2$)
  patří mezi *míry varability*. V podstatě nám říká, jaká je očekávaná druhá mocnina vzdálenosti náhodné veličiny od její střední hodnoty.

  Nechť $-\infty < \mathbb{E}X < \infty$. Potom rozptyl (pro diskrétní i spojitou náhodnou veličinu) definujeme jako
  $$\mathrm{var}X = \mathbb{E}(X - \mathbb{E}X)^2.$$

Místo tohoto vzorce se v praxi k výpočtu používá vzorce

$$\mathrm{var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Tzv. *druhý moment*, tj. $\mathbb{E}X^2$, vypočteme pro diskrétní náhodnou veličinu pomocí vzorce

$$\mathbb{E}X^2 = \sum_i x_i^2 \cdot p(x_i),$$

a pro spojitou náhodnou veličinu pomocí analogického vzorce

$$\mathbb{E}X^2 = \int_{\mathbb{R}} x^2 \cdot f(x)\mathrm{d}x.$$

Obecně, pokud chceme vypočítat střední hodnotu funkce $g$ náhodné veličiny $X$, můžeme použít následující vzorce:

$$\begin{aligned} \mathbb{E}\, g(X) &= \sum_i g(x_i) \cdot p(x_i), \qquad X \text{ diskrétní n.v.}, \\ \mathbb{E}\, g(X) &= \int_{\mathbb{R}} g(x) \cdot f(x)\mathrm{d}x, \qquad X \text{ spojitá n.v.} \end{aligned}$$

Podobně jako v případě střední hodnoty si uvedeme základní vlastnosti rozptylu. Některé další vlastnosti si uvedeme až později, neboť je pro ně nutné zavedení dalších pojmů.

VLASTNOSTI ROZPTYLU

1. $\mathrm{var}X > 0$ pro nedegenerovanou náhodnou veličinu (ale i pro degenerovanou platí $\mathrm{var}X \geq 0$),
2. $\mathrm{var}(a + bX) = b^2\, \mathrm{var}X,\ a, b \in \mathbb{R}$.

Ještě než se podíváme na další charakteristiky, uvedeme si pojem **normované náhodné veličiny**. Uvažujme nejprve veličinu $X$ se střední hodnotou $\mathbb{E}X$ a rozptylem $0 < \mathrm{var}X < \infty$. Nyní definujme náhodnou veličinu

$$X_N = \frac{X - \mathbb{E}X}{\sqrt{\mathrm{var}X}}.$$

Náhodná veličina $X_N$ má stejný typ rozdělení jako původní veličina $X$, ale má, jak čtenář může snadno nahlédnout, nulovou střední hodnotu a jednotkový rozptyl. Veličinu $X_N$ nazýváme *normovaná.*

- **Směrodatná odchylka** (značíme $\sigma$)
  je tzv. *mírou statistické disperze*, zhruba udává, jak moc je náhodná veličina odchýlena od své střední hodnoty. Je definována jako

  $$\sigma = \sqrt{\mathrm{var}X}.$$

- **Kvantily rozdělení náhodné veličiny**
  Nyní se omezíme pouze na náhodné veličiny, jejichž distribuční funkce je spojitá a rostoucí na intervalu, kde $F(x) \neq 0$ a $F(x) \neq 1$. Potom $\boldsymbol{\alpha}$**–kvantilem** (nebo též $\boldsymbol{\alpha}$**%–kvantilem**), $\alpha \in (0, 1)$, rozumíme takové číslo $u_\alpha$, že

  $$\mathbb{P}\,(X \leq u_\alpha) = \alpha.$$

  Jinými slovy řečeno je $\alpha$–kvantil takové reálné číslo, pod něž se soustředí $\alpha$ % hmoty celé pravděpodobnosti.

**Poznámka 2.** *Předpoklady uvedené pro distribuční funkci jsou nezbytné, chceme-li mít zajištěnou jednoznačnost kvantilu. Jsou-li splněny, je funkce $u : \alpha \mapsto u_\alpha$ inverzní funkce k distribuční funkci $F$. V praktických statistických úlohách se v tomto kurzu stejně jinými rozděleními než těmi, co splňují dané podmínky, zabývat nebudeme.*

NĚKTERÉ VÝZNAMNÉ KVANTILY

- $u_{0,25}$... 1. kvartil,
- $u_{0,5}$... 2. kvartil, tzv. **medián**,
- $u_{0,75}$... 3. kvartil.

### Cvičení k tomuto tématu

(i) Náhodná veličina $X$ udává počet lvů, kteří padnou při hodu čtyřmi mincemi. Stanovte pravděpodobnostní funkci $p(x_i) = \mathbb{P}(X = x_i)$ náhodné veličiny $X$, její střední hodnotu, rozptyl a směrodatnou odchylku.

(ii) Hlídač přichází do skladu náhodně vždy jednou během 24 hodin. Modelujme jeho příchody náhodnou veličinou $X$. Určete její hustotu, distribuční funkci (nakreslete také jejich grafy), $\mathbb{E}X$, $\mathrm{var}X$, $\sqrt{\mathrm{var}X}$ a 95% kvantil.

(iii) Náhodná veličina $X$ je dána distribuční funkcí $F$ v tabulce

| x | $(-\infty, -2)$ | $[-2, -1)$ | $[-1, 0)$ | $[0, 2)$ | $[2, 3)$ | $[3, +\infty)$ |
|---|---|---|---|---|---|---|
| $F(x)$ | 0 | 0,3 | 0,5 | 0,6 | 0,9 | 1 |

Nakreslete graf $F$, zrekonstrujte pravděpodobnostní funkci $p(x_i)$, a pak určete $\mathbb{E}X$, $\mathrm{var}X$, $\sqrt{\mathrm{var}X}$, $\mathbb{P}\big(X \in [-1, 2]\big)$ a $\mathbb{P}\big(1 \leq X < 6\big)$.

(iv) Náhodná veličina $X$ je zadána distribuční funkcí

$$F(x) = \begin{cases} 0 & , \quad x < 1, \\ \log x & , \quad x \in [1, 10], \\ 1 & , \quad x > 10. \end{cases}$$

Nakreslete graf $F$, stanovte hustotu $f$ náhodné veličiny $X$ (včetně jejího grafu). Dále spočtěte $\mathbb{E}X$, $\mathrm{var}X$, $\sqrt{\mathrm{var}X}$, kvartily a $\mathbb{P}\big(X > \sqrt[3]{10}\big)$.

(v) Náhodná veličina $X$ je dána hustotou

$$f(x) = \begin{cases} \frac{1}{x^2} & , \quad x \in [1, +\infty), \\ 0 & , \quad \text{jinak.} \end{cases}$$

Stanovte distribuční funkci, $\mathbb{E}X$, $\mathrm{var}X$ a všechny kvartily.

(vi) Náhodná veličina $X$ je zadána hustotou

$$f(x) = \begin{cases} \frac{1}{x^3} & , \quad x \in \big[\frac{1}{\sqrt{2}}, +\infty\big), \\ 0 & , \quad \text{jinak.} \end{cases}$$

Spočtěte $\mathbb{E}X$ a $\mathrm{var}X$.

(dcv) Náhodná veličina $X$ je dána hustotou

$$f(x) = \begin{cases} \frac{2}{3}x & , \quad x \in [0,1), \\ 1 - \frac{1}{3}x & , \quad x \in [1,3], \\ 0 & , \quad \text{jinde.} \end{cases}$$

Nakrestele graf $f$ a spočtěte

(a) $\mathbb{P}\big(X < 1\big)$, $\mathbb{P}\big(X > 1\big)$,

(b) $\mathbb{E}X$, $\mathrm{var}X$, $\sqrt{\mathrm{var}X}$,

(c) distribuční funkci náhodné veličiny $X$,

(d) kvantily $u_{0{,}1}$, $u_{0{,}5}$ a $u_{0{,}9}$.