**2. prednáška**

## Číselné charakteristiky náhodnej premennej

Často, z praktického hľadiska, je dobré zhrnúť dôležité informácie o náhodnej premennej do čísel, ktoré ilustrujú vlastnosti náhodnej premennej. Tieto čísla voláme **číselné charakteristiky** náhodnej premennej a delíme ich na:

a) **charakteristiky polohy** – vyjadrujú istý druh "stredu" rozdelenia, okolo ktorého kolíšu hodnoty náhodnej premennej (stredná hodnota, modus, medián,...),

b) **charakteristiky variability** – popisujú rozptýlenosť hodnôt náhodnej premennej okolo "stredu"(rozptyl, smerodajná odchýlka),

c) **charakteristiky, ktoré poskytujú doplňujúce údaje** o rozptýlení hodnôt okolo "stredu" (koeficient šikmosti, koeficient špicatosti).

Výpočet charakteristík je založený na <u>momentoch</u> a <u>kvantiloch</u>. Najfrekventovanejšie charakteristiky sú **momenty**, ktoré delíme na:

a) **počiatočné**

b) **centrálne**

**Počiatočný moment $k$–teho rádu** je hodnota $\nu_k$ definovaná vzťahom

$$\nu_k = \int_{-\infty}^{\infty} x^k f(x)\,\mathrm{dx},$$

ak nevlastný integrál existuje.
Najdôležitejším počiatočným momentom je počiatočný moment prvého rádu, ktorý nazývame **stredná hodnota** (*ExpectedValue*)

$$E\,(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{dx}$$

Niekedy sa označuje aj ako očakávaná hodnota alebo matematická nádej a je to číslo na reálnej osi, okolo ktorého náhodne kolíšu hodnoty náhodnej premennej.
<u>Vlastnosti strednej hodnoty:</u>
Nech $X, Y$ sú náhodné premenné, $a, b \in R$, potom platí:

a) $E\,(a) = a$

b) $E\,(aX) = a\,E\,(X)$

c) $E\,(aX + b) = a\,E\,(X) + b$

d) $E(X \pm Y) = E(X) \pm E(Y)$

e) $E(X \cdot Y) = E(X) \cdot E(Y)$, ak $X, Y$ sú nezávislé,
inak $E(X \cdot Y) = E(X) \cdot E(Y) + \operatorname{cov}(X, Y)$

**Centrálnym momentom $k$–teho rádu** rozumieme hodnotu $\mu_k$ definovaná vzťahom

$$\mu_k = \int_{-\infty}^{\infty} (X - E(X))^k f(x)\,\mathrm{dx},$$

ak nevlastný integrál existuje.
Všimnime si, že $\mu_k = E\left[(X - E(X))^k\right]$, rovnako ako $\nu_k = E\left(X^k\right)$.
Najdôležitejším centrálnym momentom je centrálny moment druhého rádu, ktorý nazývame **rozptyl** alebo **disperzia** (*Dispersion*)

$$D(X) = \int_{-\infty}^{\infty} (X - E(X))^2 f(x)\,\mathrm{dx}$$

Pre praktické výpočty je použiteľnejší vzťah $D(X) = E\left(X^2\right) - E^2(X)$. Čiže $D(X) = \nu_2 - \nu_1^2$.
Ako sme na to prišli?

Vlastnosti rozptylu:
Nech $X, Y$ sú náhodné premenné, $a, b, c \in R$, potom platí:

a) $D(a) = 0$

b) $D(aX) = a^2 D(X)$

c) $D(X \pm Y) = D(X) + D(Y)$, ak $X, Y$ sú nezávislé

d) $D(X \pm Y) = D(X) \pm 2\operatorname{cov}(X, Y) + D(Y)$, ak $X, Y$ sú závislé

e) $D(aX + bY + c) = a^2 D(X) + 2ab\operatorname{cov}(X, Y) + b^2 D(Y)$

Hlavne v štatistike sa budeme stretávať s pojmom normovaná náhodná premenná. Aká premenná má tento prívlastok?
Náhodná premenná $Z$ je **normovaná**, ak pre ňu platí: $E(Z) = 0$ a $D(Z) = 1$.
Ak náhodná premenná $X$ má strednú hodnotu $E(X)$ a $D(X) \neq 0$, potom náhodná premenná

$$Z = \frac{X - E(X)}{\sqrt{D(X)}}$$

je normovaná náhodná premenná.

Ukážeme, že je to pravda:

Podiel centrálneho momentu tretieho rádu a tretej mocniny smerodajnej odchýlky náhodnej premennej $X$ sa nazýva **koeficient asymetrie** alebo **koeficient šikmosti (Skew)**:

$$\alpha_1 = \frac{\mu_3}{\sigma^3}$$

Koeficient šikmosti je mierou symetrie rozdelenia pravdepodobnosti náhodnej premennej $X$. Symetrické rozdelenia majú koeficient šikmosti rovný nule, nesymetrické rozdelenia majú pozitívne alebo negatívne zošikmenie. Vyjadrené matematicky:

- ak $\alpha_1 = 0$, rozdelenie je symetrické,
- ak $\alpha_1 < 0$, rozdelenie je pozitívne zošikmené,
- ak $\alpha_1 > 0$, rozdelenie je negatívne zošikmené.

Koeficient **špicatosti (Kurtosis)** označíme $\alpha_2$ a definujeme:

$$\alpha_2 = \frac{\mu_4}{\sigma^4} - 3$$

Normálne rozdelenie má koeficient špicatosti rovný nule. Kladné hodnoty koeficienta špicatosti znamenajú, že rozdelenie pravdepodobnosti náhodnej premennej je v porovnaní s normálnym rozdelením strmšie (špicatejšie), záporné hodnoty znamenajú, že rozdelenie je menej strmé ako normálne rozdelnie.

V štatistike sa budeme stretávať s číselnými hodnotami, ktoré sa nazývajú **kvantily**. Pre číslo $\alpha \in (0,1)$ zavedieme $\alpha$-kvantil pomocou distribučnej funkcie spojitej náhodnej premennej. **$\alpha$-kvantil** je hodnota $x_\alpha$, pre ktorú platí:

$$F(x_\alpha) = \alpha$$

$\alpha$-kvantil je hodnota, ktorá delí plochu pod grafom hustoty pravdepodobnosti v pomere $\alpha : (1-\alpha)$. Pokiaľ je distribučná funkcia rastúca (a teda aj prostá ), $\alpha$-kvantil je možné vyjadriť v tvare:

$$x_\alpha = F^{-1}(\alpha)$$

Niektoré kvantily majú špeciálne názvy:

| **Hodnota $\alpha$** | **Kvantil** | **Názov** |
|---|---|---|
| $0,5$ | $x_{0,5}$ | medián |
| $0,25$ | $x_{0,25}$ | dolný kvartil |
| $0,75$ | $x_{0,75}$ | horný kvartil |
| $0,1; 0,2; \ldots; 0,9$ | $x_{0,1}; x_{0,2}; \ldots; x_{0,9}$ | decily |
| $0,01; 0,02; \ldots; 0,99$ | $x_{0,01}; x_{0,02}; \ldots; x_{0,99}$ | percentily |

**Medián**, ozn. $m_e$, je teda hodnota, ktorá rozdelí interval možných hodnôt náhodnej premennej na dva rovnako pravdepodobnostné intervaly, t.j.
$P(X < m_e) = P(X \geq m_e)$.
**Modus**, ozn. $m_o$, je hodnota, v ktorej má hustota pravdepodobnosti lokálne maximum.