Štatistika 1
Miery variability

2.prednáška

Osnova prednášky:

- (1) Členenie mier variability
- (2) Absolútne miery variability
- (3) Relatívne miery variability
- (4) Rozklad rozptylu na zložky
- (5) Kovariancia
- (6) Koeficient korelácie

(1)Členenie mier variability

Variabilita

je premenlivosť hodnôt znaku alebo navzájom alebo voči určitej typickej konštante.

Hľadiská členenia mier variability

- vplyv hodnôt znaku na veľkosť charakteristiky
  - charakteristiky, ktoré **nie sú ovplyvnené každou hodnotou znaku** (variačné rozpätie, kvantilové rozpätie, kvantilová odchýlka)
  - charakteristiky variability, ktoré **sú ovplyvnené každou hodnotou znaku** (priemerná odchýlka, smerodajná odchýlka, rozptyl).

- hľadisko interpretácie
  - **absolútne miery variability –** miery v pôvodných merných jednotkách, prípadne v ich štvorcoch (variačné rozpätie, kvantilové rozpätie, kvantilová odchýlka, priemerná odchýlka, rozptyl, štandardná odchýlka),
  - **relatívne (pomerné) miery** variability - v percentách (variačný koeficient (V_k) a pomerná priemerná odchýlka (D_x)).

(2) Absolútne miery variability

- **Variačné rozpätie**

$$R = x_{max} - x_{min}$$

  - rozdiel najväčšej a najmenšej hodnoty
  - závisí len od extrémnych hodnôt, neinformuje o skutočnej variabilite medzi extrémami

- **Kvantilové rozpätie** $R_Q = Q^{\alpha}_{\alpha-1} - Q^{\alpha}_{1}$
  - rozdiel medzi horným a dolným kvantilom
  - kvartilové rozpätie $R_Q = Q^4_3 - Q^4_1$

Kvantily

- rozdeľujú rad hodnôt znakov, usporiadaný podľa veľkosti, na určitý počet skupín (α) s rovnakým počtom prvkov.
- medián ($\alpha = 2$)

  kvartily ($\alpha = 4$)

  decily ($\alpha = 10$)

  percentily ($\alpha = 100$)

- **Kvantilová odchýlka**

  je definovaná ako aritmetický priemer kladných odchýlok susedných kvantilov

$$Q^{\alpha} = \frac{\left(Q_{\alpha-1}^{\alpha} - Q_{\alpha-2}^{\alpha}\right) + \left(Q_{\alpha-2}^{\alpha} - Q_{\alpha-3}^{\alpha}\right) + \ldots + \left(Q_{2}^{\alpha} - Q_{1}^{\alpha}\right)}{\alpha - 2} = \frac{Q_{\alpha-1}^{\alpha} - Q_{1}^{\alpha}}{\alpha - 2}$$

- **Kvartilová odchýlka** je polovičné rozpätie medzi horným a dolným kvartilom

$$Q = \frac{Q_3^4 - Q_1^4}{2}$$

Priemerná odchýlka

- je aritmetickým priemerom absolútnych odchýlok jednotlivých hodnôt sledovaného znaku od strednej hodnoty
- podľa toho, od ktorej strednej hodnoty meriame jednotlivé odchýlky, dostávame
  - priemerná odchýlka vzhľadom na aritmetický priemer
  - priemerná odchýlka vzhľadom na medián.

$$d_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{m} \left|x_i - \bar{x}\right| \cdot n_i \qquad d_{\tilde{x}} = \frac{1}{n}\sum_{i=1}^{n} \left|x_i - \tilde{x}\right|$$

Rozptyl (variancia)

- je aritmetický priemer zo štvorcov odchýlok hodnôt znaku od aritmetického priemeru

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{m}(x_i - \bar{x})^2 n_i$$

- výpočtový tvar

$$s_x^2 = \overline{x^2} - \bar{x}^2 \qquad \text{kde} \quad \overline{x^2} = \frac{\sum_{i=1}^{m} x_i^2 n_i}{n}$$

Smerodajná odchýlka

$$s_x = \sqrt{s_x^2}$$

- interpretuje sa ako $\pm$ odchýlka od aritmetického priemeru

Vlastnosti rozptylu

- Rozptyl konštanty sa rovná nule.
- Ak ku každej hodnote znaku pripočítame konštantu, rozptyl sa nezmení.
- Ak každú hodnotu znaku vynásobíme nenulovou konštantou, rozptyl sa zmení súčinom štvorca tejto konštanty.
- Rozptyl súčtu (rozdielu) dvoch znakov sa rovná súčtu (rozdielu) rozptylov týchto znakov zväčšenému (zmenšenému) o dvojnásobok kovariancie

(3) Pomerné miery variability

- Variačný koeficient

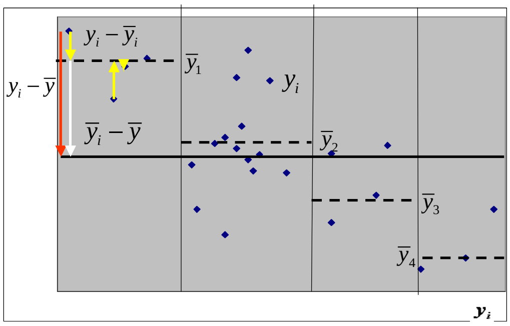
$$V_k = \frac{s_x}{\bar{x}} \cdot 100\ [\%]$$

- Pomerná priemerná odchýlka

$$D_x = \frac{\bar{d}}{\bar{x}} \cdot 100\ [\%]$$

Používajú sa na porovnanie variability znakov, ktoré sa líšia svojou úrovňou alebo mernými jednotkami.

(4) Rozklad rozptylu na zložky



celkový rozptyl =
vnútroskupinový rozptyl + medziskupinový rozptyl

$$S_y^2 = \overline{S_i^2} + S_{\bar{y}}^2 = \frac{\sum_{i=1}^{m} S_i^2 \cdot n_i}{\sum_{i=1}^{m} n_i} + \frac{\sum_{i=1}^{m} \left(\bar{y}_i - \bar{y}\right)^2 \cdot n_i}{\sum_{i=1}^{m} n_i}$$

Kovariancia

- je aritmetický priemer súčinov odchýliek hodnôt znakov od ich aritmetických priemerov
- charakterizuje typ lineárnej závislosti

v jednoduchom tvare:

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}(x_i-\overline{x})\cdot(y_i-\overline{y})}{n}=\overline{xy}-\overline{x}\cdot\overline{y}$$

- kovariancia vo váženom tvare

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i-\overline{x})\cdot(y_j-\overline{y})\cdot n_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{s}n_{ij}}=$$

$$=\frac{\sum_{i=1}^{r}\sum_{j=1}^{s}x_i\cdot y_j\cdot n_{ij}}{n}-\frac{\sum_{i=1}^{r}x_i\cdot n_{i.}}{n}\cdot\frac{\sum_{j=1}^{s}y_j\cdot n_{.j}}{n}$$

Interpretácia kovariancie

Ak medzi znakmi X a Y predpokladáme lineárny vzťah

- cov(x,y) > 0 – medzi znakmi je priama zavislosť
- cov(x,y) < 0 – znaky sú nepriamo závislé
- cov(x,y) = 0 – znaky sú nekorelované

Koeficient korelácie

- vzťah: $$r(x,y)=\frac{\operatorname{cov}(x,y)}{s_x s_y}$$
- môže nadobudnúť hodnoty z intervalu $\langle -1;1\rangle$
- meria intenzitu lineárnej závislosti
- čím je hodnota koeficientu korelácie bližšia k ± 1, tým je závislosť silnejšia

Ďakujem za pozornosť.