

III. Štatistické výpočty

Úvod.

Čo je to štatistika?

Dve základné úrovne:

opisná (deskriptívna) štatistika

Inferenčná (matematická) štatistika

Rozdelenie z hľadiska dimenzie:

jednorozmerný dátový súbor

viacrozmerný dátový súbor

Kap. 1 Opisná štatistika jednorozmerného štatistického súboru

Definícia dátového (štatistického) súboru:

Dátový súbor je postupnosť prvkov, v ktorej nezáleží na ich poradí.

Definičný obor – objekty.

Obor hodnôt – znak.

Povaha znaku: kvalitatívny – nominálny

– ordinálny

kvantitatívny

Príklady.

1. Prieskum volebných preferencií.

2. Abstraktný konečný dátový súbor.

Sumarizácia – kategorizácia.

Tabuľky, diagramy, histogramy

Číselné charakteristiky dátového súboru

Charakteristiky polohy

Aritmetický priemer

Dátový súbor $x = (x_i)_{i=1,2,\dots,n}$

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

Hodnoty znakov $y = (y_j)_{j=1,2,\dots,k}$,

početnosti znakov $f = (f_j)_{j=1,2,\dots,k}$,

$$\bar{x} = \frac{1}{n} \sum_{j=1}^k f_j y_j$$

Kategorizovaný súbor, stredy kategórií $s = (s_j)_{j=1,2,\dots,k}$,
početnosti kategórií $f = (f_j)_{j=1,2,\dots,k}$,

$$\bar{y} = \frac{1}{n} \sum_{j=1}^k f_j s_j$$

Medián m

Najmenej polovica dátového súboru je $\leq m$ a najmenej polovica dátového súboru je $\geq m$.

Ak je rozsah n súboru párny, $m = (x_{n/2} + x_{n/2+1}) / 2$

Ak je rozsah n súboru nepárny, $m = x_{(n+1)/2}$

Medián je robustnou charakteristikou polohy.

p-kvantil q_p

Najmenej 100p% dátového súboru je $\leq q_p$ a najmenej 100(1-p)% dátového súboru je $\geq m$.

Ak je $n \cdot p$ celé, $q_p = (x_{n \cdot p} + x_{n \cdot p + 1}) / 2$,

inak $q_p = x_{[n \cdot p] + 1}$.

0,25-kvantil sa nazýva **dolný (prvý) kvartil** q_1

0,75–kvantil sa nazýva **horný (tretí) kvartil** q_3

Charakteristiky rozptylu

Dátové rozpätie $R = \max(x) - \min(x)$

Medzikvartilové rozpätie $mkr = q_3 - q_1$

Výberový rozptyl

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

Výberová smerodajná odchýlka $s = \sqrt{s^2}$.