

Statistika

- Cíle:
1. Chápat pojmy statistický soubor, rozsah souboru, statistická jednotka, statistický znak, umět sestavit tabulku rozdělení četností, umět znázornit spojnicový diagram a sloupcový diagram / kruhový diagram /.
 2. Z charakteristik polohy určit a vhodně použít typy průměrů.
 3. Z charakteristik variability určit rozptyl, absolutní odchylku, směrodatnou odchylku a variační koeficient.

Statistika je vědní obor zabývající se zkoumáním jevů, které mají hromadný charakter.

Základy popisné statistiky

Pojem statistika slouží k označení : - statistických údajů a jejich funkcí

- statistické činnosti
- statistické teorie
- matematické statistiky

Základní statistické pojmy

Statistický soubor je neprázdná konečná množina objektů, které mají společné vlastnosti.

Rozsah souboru n je počet všech prvků množiny.

Statistické jednotky nebo prvky statistického souboru nazýváme prvky množiny.

Statistické údaje / data / jsou kvantitativní údaje zjištěné zkoumáním hromadných jevů.

Statistický znak je společná vlastnost statistických jednotek / značí se x /.

Hodnoty znaku – jednotlivé údaje znaku – označíme $x_1, x_2 \dots x_3$

Dělení znaků - kvantitativní a kvalitativní

Zjišťování hodnot volených znaků v určitém statistickém souboru se nazývá **statistické šetření**.

Rozdělení četnosti

Počet statistických jednotek, jimž přísluší stejná hodnota znaku se nazývá **absolutní četnost hodnoty znaku x_i označené n_i** .

Podíl absolutní četnosti znaku a rozsahu souboru se nazývá **relativní četnost s označením v_i** .

Součet absolutních četností je roven rozsahu souboru, součet relativních četností je roven 1.

Tabulka rozdělení četností

i	x_i	n_i	v_i	v_i (%)
1				
2				
.				
.				

Skupinové – intervalové rozdělení četnosti

Je-li rozsah statistického souboru velký je výhodné blízké hodnoty uspořádat do skupin, intervalů, které by byly charakterizovány středem intervalu. Počet těchto intervalů by měl odpovídat rozsahu souboru. Využijeme jedno z pravidel – Sturgesovo pravidlo

$$k = 1 + 3,3 \log n$$

k – počet intervalů

n – rozsah souboru

Vzorová úloha : V podniku je 1 000 pracovníků, jejichž příjem je od 5 000,- do 25 000,- Kč.
Navrhněte vhodný počet intervalů a formu intervalového rozdělení.

Řešení: počet intervalů $k = 1 + 3,3 \log 1\,000$

$$k = 10,9 = 11$$

intervaly $(25\,000 - 5\,000) / 11 = 1818$ Kč

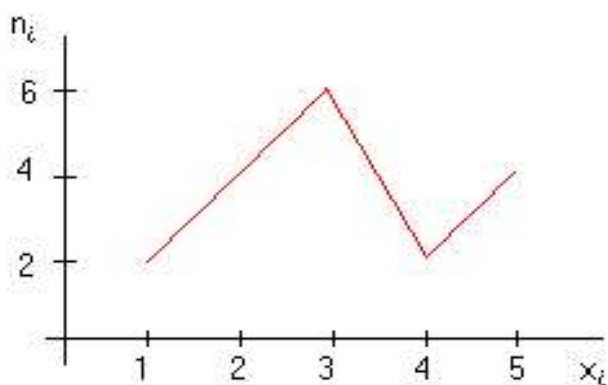
Tabulka skupinového rozdělení

i	interval	střed intervalu x_i	n_i
1			
2			
·			
·			
·			

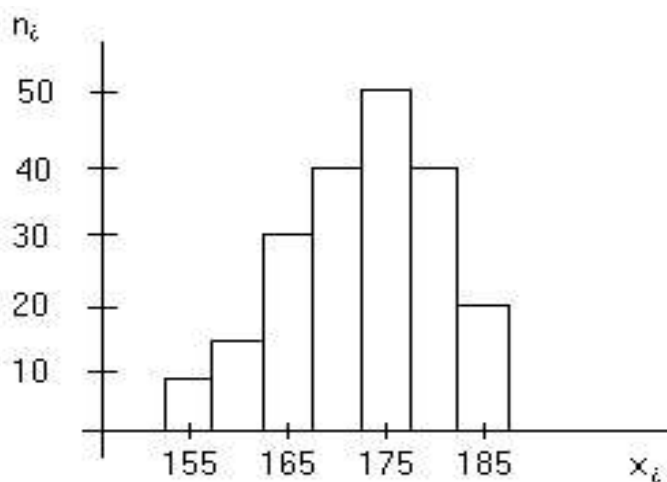
Grafické znázornění

Jedná se o závislost absolutní četnosti /relativní/ na hodnotě znaku.

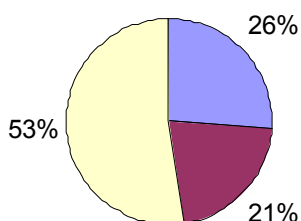
1. **Spojnicový diagram** / polygon četnosti /



2. **Sloupcový diagram** / histogram četnosti /



3. **Kruhový diagram** / hodnoty znaku jsou znázorněny kruhovými výsečemi, jejichž obsahy jsou přímo úměrné relativní četnosti v % /.



Charakteristiky znaku statistického souboru

Jedná se o čísla, která podávají stručnou souhrnou informaci o uvažovaném statistickém souboru z různých hledisek. Jedná-li se o kvantitativní znak jde především o charakteristiky polohy / úrovně / charakteristiky variability / proměnnosti /.

Charakteristiky polohy – střední hodnoty

Jedná se o čísla, charakterizující „průměrnou hodnotu“ sledovaného kvantitativního znaku.

aritmetický průměr je dán podílem součtu hodnot znaků a rozsahu souboru

- pro větší rozsah užíváme **vážený aritmetický průměr**

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

harmonický průměr nenulových hodnot statistického znaku je definován jako podíl rozsahu souboru a součtu převrácených hodnot znaku

- hodnoty rovnoměrně vztažené kolem průměru

$$x_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_j}}$$

geometrický průměr z kladných hodnot znaku je definován jako n-tá odmocnina ze součinu hodnot znaku

- četnosti – skutečné objemy výroby

$$x_G = \sqrt[n]{x_1 \dots x_n}$$

modus je nejčastěji se vyskytující hodnota mezi znaky, značíme **mod(x)**

medián je prostřední člen mezi znaky, jestliže je uspořádáme podle velikosti / lichý, sudý počet znaků /, značíme **med(x)**

Charakteristiky variability – odlišnost hodnot příslušných znaků

variační rozpětí je pouze orientační charakteristika a určuje rozdíl mezi největší a nejmenší hodnotou znaku

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

průměrná absolutní odchylka je dokonalejší charakteristikou a určuje aritmetický průměr absolutních hodnot odchylek znaku všech prvků souboru od aritmetického průměru hodnot znaku

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

rozptyl je nejpoužívanější charakteristikou a určuje průměrnou čtvercovou odchylku od aritmetického průměru

- při uspořádání údajů do tabulky rozdělení četností používáme váženou formu rozptylu

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

$$Sx^2 = \frac{\sum_{i=1}^n (x_j - \bar{x})^2 n_j}{\sum_{i=1}^n n_j}$$

směrodatná odchylka je blízká průměrné odchylce, čím je směrodatná odchylka menší, tím blíže jsou hodnoty znaku kolem aritmetického průměru

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

variační koeficient je relativní mírou variability, má smysl tehdy, nabývá-li znak nezáporné hodnoty, je dán podílem směrodatné odchylky a aritmetického průměru, výsledek uvádíme v procentech

$$v_x = \frac{s_x}{x}$$

Na závěr ukázka samostatných prací studentů při zpracování statistického souboru.