

## Nové možnosti rozvoje vzdělávání na Technické univerzitě v Liberci

Specifický cíl A3: Tvorba nových profesně zaměřených studijních programů

NPO\_TUL\_MSMT-16598/2022



# Základní statistické pojmy

Ing. Vladimíra Hovorková Valentová, Ph.D.

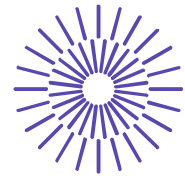


Financováno  
Evropskou unií  
NextGenerationEU



Národní  
plán  
obnovy

MSMT  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



## Základní statistické pojmy

### **Statistika**

- tento pojem lze chápat v zásadě ve třech pojetích:
  - 1) číselné nebo slovní údaje (data) a jejich souhrny o hromadných jevech;
  - 2) praktická činnost spočívající ve sběru, zpracování a vyhodnocování dat o hromadných jevech;
  - 3) teoretická disciplína (věda), která zkoumá zákonitosti hromadných jevů, resp. souhrn vědeckých metod sběru, zpracování a analyzování dat

V každé definici statistiky je uvedeno, že se zabývá hromadnými jevy.

### **Hromadné jevy**

- takové skutečnosti, které se vyskytují mnohokrát a mohou se znovu opakovat;
- jevy, které se vyskytují v masovém měřítku u velkého počtu prvků.

### **Statistický soubor**

- souhrn všech jednotek, které jsou předmětem zkoumání.
- množina prvků s přesně stanovenými shodnými vlastnostmi (např. množina osob, organizací, strojů atd.).

### **Statistická jednotka**

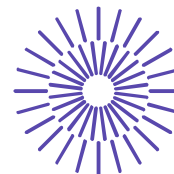
- prvek statistického souboru;
- individuální nositel vlastností daného statistického souboru.

### **Rozsah statistického souboru**

- počet jednotek statistického souboru (symbolické značení – ***n***, ***N***).

! Existují dvě možnosti přístupu ke statistickému souboru – jejich chápání je relativní.

- 1) Základní soubor (populace) = statistický soubor všech jednotek, které jsou předmětem zkoumání, obvykle velmi rozsáhlý, rozsah značíme ***N***.
- 2) Výběrový soubor (výběr) = vzorek ze základního souboru, pořízený tak, že se určitým způsobem vyberou pouze některé jednotky, rozsah značíme ***n***.



### **Statistický znak**

- označení (odraz) určité vlastnosti, kterou má v té či oné míře každá jednotka daného statistického souboru;
- u souboru osob např. věk, váha, výška, atd.
- ekvivalentní označení: statistická proměnná.

### **Hodnota statistického znaku (= pozorování)**

- míra dané vlastnosti (statistického znaku) u každé jednotky statistického souboru.

*! Počet hodnot (pozorování) = rozsah souboru.*

### **Obměna (= varianta) statistického znaku**

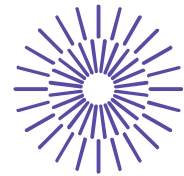
- hodnota ve smyslu vyjádření různého stupně dané vlastnosti.

*! Počet variant  $\leq$  rozsah souboru.*

**Statistický znak shodný:** v daném statistickém souboru nabývá pouze jedné varianty.

**Statistický znak proměnný:** v daném statistickém souboru nabývá více než jedné varianty.

Ekvivalentní označení = **statistická proměnná.**



## Zpracování dat

- Ke statistickému zkoumání jsou třeba *hodnoty proměnných = data = údaje*; získáme je statistickým šetřením.
- Tato data pak jsou zpracována a vyhodnocena.
- Většinou jde o velké množství údajů, které jsou značně nepřehledné.

Existují 3 základní způsoby zpracování dat:

1) Tabulka.

2) Graf.

3) Popisné charakteristiky.

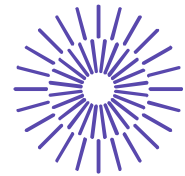
- Prvním krokem je proto zpřehlednění (setřídění) dat formou *tabulek a grafů*.
- Cílem je, aby vynikly charakteristické rysy a zákonitosti analyzovaného souboru.

### Tabulka prostého rozdělení četností

Odměna proměnné $x_i$	Četnost		Kumulativní četnost	
	absolutní $n_i$	relativní $p_i$	absolutní	relativní
$x_1$	$n_1$	$p_1$	$n_1$	$p_1$
$x_2$	$n_2$	$p_2$	$n_1 + n_2$	$p_1 + p_2$
$x_3$	$n_3$	$p_3$	$n_1 + n_2 + n_3$	$p_1 + p_2 + p_3$
.	.	.	.	.
.	.	.	.	.
$x_k$	$n_k$	$p_k$	$n$	1
Celkem	$n$	1	×	×

$$p_i = \frac{n_i}{\sum_{i=1}^k n_i} = \frac{n_i}{n} ; \sum_{i=1}^k n_i = n ; \sum_{i=1}^k p_i = 1$$

- Je to *tabulka prostého rozdělení četností*.
- Takováto tabulka je výsledkem zpracování diskrétní proměnné s několika málo odměnami.
- Statistický znak, který jsme označili  $x$ , nabývá odměn  $x_i, i = 1, 2, \dots, k$ , tj.  $x_1, x_2, \dots, x_k$ .
- Odměny  $x_i$  se nejčastěji uspořádávají do tabulky rozdělení četností VZESTUPNĚ.
- V případě zpracování diskrétní proměnné s mnoha odměnami nebo spojitě proměnné není použitelná, pak je třeba data třídit do tabulky *intervalového rozdělení četností*.



### **Tabulka intervalového rozdělení četností**

- Je třeba stanovit optimální počet intervalů ( $k$ ), na které rozdělíme variační rozpětí ( $R$ ):  
$$R = x_{max} - x_{min} .$$
- K tomu slouží různá pravidla (např. Sturgesovo pravidlo:  $k \approx 1 + 3,3 \cdot \log n$ ).
- Každý interval lze při výpočtech zastoupit jeho středem (výsledky takovýchto výpočtů jsou přibližné).
- Nejlepším reprezentantem každého intervalu je skupinový průměr, tj. aritmetický průměr hodnot v dané skupině.

#### *Zásady při konstrukci intervalů:*

- Všechny intervaly by měly být stejně široké.
- Intervaly konstruujeme tak, aby střed intervalu bylo celé číslo.
- Hranice intervalů označujeme tak, aby bylo možné zařadit každou jednotku.

## **Grafy**

Existuje mnoho druhů grafů, je třeba vybrat takový, který odpovídá charakteru dat.

### **Polygon četností**

- graf vhodný pro znázornění prostého rozdělení četností;
- spojnicový graf.

### **Histogram četností**

- graf vhodný pro znázornění intervalového rozdělení četností;
- sloupkový graf.

### **Výsečový graf (piechart), sloupkový graf (barchart)**

- grafy vhodné pro znázornění rozdělení četností nominální proměnné.