

C. STATISTIKA

1. Statistický soubor, jednotka, znak, četnost

STATISTIKA – věda o metodách sběru, zpracování a vyhodnocování statistických údajů

- zkoumá společenské, přírodní, technické aj. jevy **na dostatečně rozsáhlém souboru údajů**
- **matematická statistika** – vychází ze shromážděných statistických údajů, zabývá se jejich matematickým zpracováním a rozбором výsledků
- **využití statistiky**
 - reklamní agentury (výzkumy chování, dopad kampaní,...)
 - politika (výzkumy veřejného mínění, přizpůsobování politiky,...)
 - média (sledovanost, poslechovost,...)
 - výzkum (účinnost léků,...)

ZÁKLADNÍ POJMY

- **statistické údaje (data)** – číselné údaje o společenských, přírodních, technických aj. skutečnostech, o tzv. hromadných jevech
- **statistický soubor** – soubor (množina) osob, věcí, událostí, jevů apod. shromážděných na základě toho, že mají určité společné vlastnosti
 - **základní**: soubor všech vybraných jednotek
 - **výběrový**: obsahuje jen část jednotek základního souboru, pracujeme s ním zpravidla tehdy, pokud jsou základní soubory příliš rozsáhlé a zkoumání všech jednotek by bylo neproveditelné, příliš pracné, časově náročné nebo nákladné
- **statistické jednotky** – jednotlivé prvky statistického souboru (např. žáci třídy)
- **rozsah souboru (zn. n)** – počet všech jednotek (prvků) statistického souboru (např. počet žáků třídy)
- **statistický znak (zn. x)** – společná vlastnost prvků statistického souboru, jejíž proměnnost je předmětem statistického zkoumání (např. známka z matematiky)
- **hodnoty znaku (zn. $x_1, x_2, \dots, x_n$)**
 - **kvantitativní**: hodnoty vyjádřené čísly (např. známka z testu, výška, hmotnost, hrubý roční příjem, hektarový výnos atd.)
 - **kvalitativní**: hodnoty vyjádření slovním popisem
 - **alternativní**: dány určitým jevem a jeho opakem (např. žena – muž, prospěl – neprospěl, voják – nevoják)
 - **znaky nabízející více možností hodnot** (např. národnost, barva očí, náboženství)

Příklady

- ① Statistika dopravních nehod na silnicích ČR v 1. čtvrtletí roku 2011

statistický soubor: všechny nehody ve sledovaném čtvrtletí

statistické znaky: příčina nehody (kvalitativní)

možné **hodnoty** (předem vymezeny, jedna z hodnot musí vždy nastat): nepřiměřená rychlost, nedání přednosti v jízdě, nesprávné předjíždění, špatný technický stav vozidla, vliv alkoholu, jiné příčiny (méně obvyklé příčiny shrnuty v jednu hodnotu),

příčina nezjištěna (některé nehody mohou mít víc než jednu příčinu – např. nepřiměřenou rychlost a vliv alkoholu – aby byl dodržen požadavek neslučitelnosti volíme za příčinu hlavní z nich podle nějakého předem stanoveného pravidla)

hmotná škoda: *kvantitativní znak*

zranění osob: *kvalitativní alternativní*, lze např. doplnit kvantitativním znakem – počet zraněných osob

- ② V jisté třídě byly zjišťovány pro účely výuky statistiky zjišťovány mimo jiné tyto znaky: výška, kuřáctví, oblíbený předmět, průměrný čistý měsíční příjem, známka z matematiky. Rozdělte je na kvantitativní a kvalitativní. Které z nich jsou alternativní?

Kvalitativní znaky: kuřáctví, oblíbený předmět

– alternativní znaky: kuřáctví

Kvantitativní znaky: výška, příjem, známka

Při statistickém šetření se zpravidla vyšetřuje řada znaků, které nás zajímají jak každý zvlášť, tak i ve vzájemném vztahu. My se zatím omezíme na případy, kdy nás zajímá jediný znak.

2. Četnost, rozdělení četností a grafické znázornění

Předpoklady:

- zajímá nás jediný statistický znak x , který nabývá hodnot $x_1, x_2, \dots, x_n$
- rozsah souboru je n , tj. počet všech jednotek (prvků statistického souboru)
- často znak x nabývá jen určitého počtu r různých hodnot, ozn. $x_1^*, x_2^*, \dots, x_r^*$
- počet různých hodnot znaku r (ve většině šetření) je menší než počet jednotek souboru n ($r \leq n$), tj. že několik různých jednotek téhož souboru nabývá stejných hodnot

ABSOLUTNÍ ČETNOST (ČETNOST) hodnoty znaku x_j^* – zn. n_j

- udává počet statistických jednotek, kterým přísluší stejná hodnota znaku x_j^* , tj. kolikrát se hodnota znaku x_j^* vyskytla mezi hodnotami $x_1, x_2, \dots, x_n$
- součet četností všech možných hodnot znaku se rovná počtu všech jednotek (rozsahu) souboru n :

$$\sum_{j=1}^r n_j = n$$

RELATIVNÍ ČETNOST – zn. v_j

- podíl absolutní četnosti n_j hodnoty znaku x_j^* a rozsahu souboru n (značí, jaká část souboru má hodnotu znaku x_j^*): $v_j = \frac{n_j}{n}$
- součet všech relativních četností se rovná jedné (lze i v procentech – pak 100 %)

$$\sum_{j=1}^r v_j = 1$$

TABULKA ROZDĚLENÍ ČETNOSTÍ A RELATIVNÍCH ČETNOSTÍ

(nebo jen tabulka rozdělení četností, tabulka rozdělení relativních četností)

znak x	x_1^*	x_2^*	...	x_r^*
četnost n_j	n_1	n_2	...	n_r
relativní četnost v_j	v_1	v_2	...	v_r

Příklady

- ① Ze seznamu 143 pracovníků podniku s údaji o počtu rodinných příslušníků bylo získáno následující rozdělení četností. Vypočítejte relativní četnosti a vyjádřete je i v procentech.

počet rodin. přísluš.	1	2	3	4	5	6	7	8
četnost	7	29	36	42	21	4	3	1
relativní četnost	0,0489	0,2027	0,2517	0,2937	0,1470	0,0280	0,0210	0,007
relativní četnost v %	4,89	20,27	25,17	29,37	14,70	2,80	2,10	0,7

- ② Ve třídě je 32 studentů, z nichž dva studenti mají z testu známku 1, deset studentů známku 2, dvanáct studentů známku 3, šest studentů známku 4 a dva studenti známku 5. Zpracujte tabulku rozdělení četností a relativních četností, přičemž sledovanou hodnotou znaku je známka z testu.

$n = 32$

známka	1	2	3	4	5
četnost	2	10	12	6	2
relativní četnost	0,0625	0,3125	0,375	0,1875	0,0625
relativní četnost v procentech	6,25	31,25	37,5	18,75	6,25

- ③ Pokud je rozsah statistického souboru velký a postupují-li hodnoty znaku po příliš malých krocích, pak se blízké hodnoty znaku sdružují do skupin tvořených intervaly (nejjednodušší je zvolit všechny intervaly stejné šířky, hodnoty z téhož intervalu pak zaokrouhlujeme na střed intervalu).
Tabulku rozdělení četností při měření výšky po 1 cm (hodnoty sdružujeme v intervaly např. po 5 cm, pak zapisujeme jedním ze dvou způsobů:

$n = 194$

x_j^*	158–162	163–167	168–172	173–177	178–182	183–187	188–192
n_j	9	20	30	82	35	14	4

x_j^*	160	165	170	175	180	185	190
n_j	9	20	30	82	35	14	4

- ④ Souborem je 320 žáků školy, znakem je druhý volitelný jazyk: angličtina, němčina, ruština. Vypočítejte relativní četnost v procentech.

jazyk	A	N	R
četnost	176	105	39
relativní četnost v %	55	32,81	12,19

GRAFICKÉ ZNÁZORNĚNÍ ROZDĚLENÍ ČETNOSTÍ

- zpravidla pracujeme s grafy v pravoúhlé soustavě souřadnic, na vodorovnou osu (x) vynášíme hodnoty znaku nebo intervalu, na svislou osu (y) vynášíme jim odpovídající četnosti nebo relativní četnosti
- nejčastěji rozdělení četností znázorňujeme polygonem četností a histogramem četností, u kvalitativních znaků také kruhovým diagramem

Polygon četnosti (spojnicový diagram)

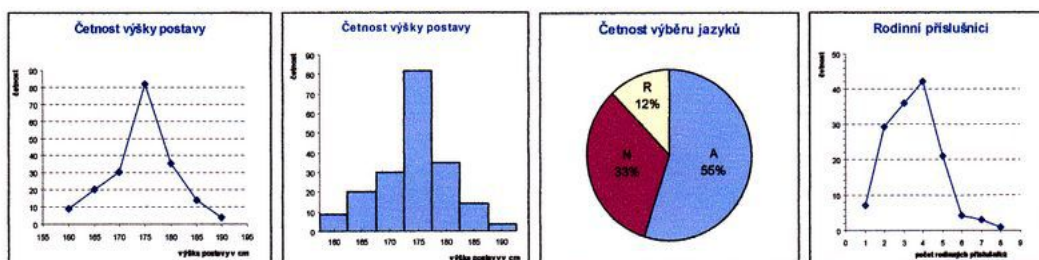
- získáme spojením bodů, jejichž první souřadnice (x) je hodnota kvantitativního znaku (intervalu) a druhá souřadnice (y) je odpovídající četnost

Sloupkový diagram (histogram)

- používá se, pokud jsou hodnoty grafu sdruženy v intervaly, tyto intervaly tvoří základny sloupků, odpovídající četnosti udávají jejich výšky (kdyby intervaly nebyly stejně dlouhé, musela by být četnostem rovna nikoli výška, ale plošný obsah sloupků)

Kruhový diagram

- používá se pro rozdělení četností kvalitativního znaku, různým hodnotám znaku odpovídají kruhové výseče, jejichž plošné obsahy jsou úměrné četnostem)



- ⑤ Sestav tabulku četností a relativních četností všech známek z matematiky. Zkontroluj součty.

Známky jednotlivých studentů: 3, 2, 2, 4, 3, 2, 3, 3, 3, 3, 4, 3, 2, 3, 4, 2, 3, 1

$n = 19$

známka	1	2	3	4	5
četnost	1	5	10	3	0
relativní četnost	0,053	0,263	0,526	0,158	0
relativní četnost v procentech	5,3	26,3	52,6	15,8	0

- ⑥ Sestrojte pro příklad 4 sloupkový diagram pro rozdělení četností.

