8. Základy statistiky

Statistika je vědní obor, který se zabývá zpracováním hromadných jevů. Tvoří základ pro řadu procesů řízení, rozhodování a organizování, protože na základě použití řady statistických metod nám přibližuje zkoumaný jev a zákonitosti s ním spojené.

8.1 Statistický soubor

Příklad 1 : V tabulce jsou uvedeny údaje o žácích na základních školách k 30. září 1998

	Počet žáků	V procentech	Počet dívek
V 1. až 9. ročníku celkem	1 082 415	100	527 455
Z toho v 1. ročníku	123 221	11,4	59 494
V 2. ročníku	128 384	11,9	62 879
Ve 3. ročníku	123 804	11,4	60 579
Ve 4. ročníku	125 407	11,6	61 581
V 5. ročníku	125 972	11,6	61 544
V 6. ročníku	116 563	10,8	56 473
V 7. ročníku	116 376	10,8	56 239
V 8. ročníku	113 026	10,4	54 688
V 9. ročníku	109 662	10,1	53 981

Tabulka byla sestavena na základě **statistického šetření**. Pozorně si ji prohlédni a pak zodpověz těchto deset otázek:
a) Kolik žáků bylo 30. 9. 1998 v 1. až 9. ročníku celkem?
b) Kolik z nich bylo k tomuto datu v 8. ročníku?
c) Ve kterém ročníku bylo nejvíce žáků? A ve kterém nejméně?
d) Bylo více žáků v 1. až 5. ročníku, nebo v 6. až 9. ročníku?
e) Které číslo tvoří **základ** pro výpočet počtů procent žáků v jednotlivých ročnících?
f) Jsou počty procent vypočítány správně? Zkontrolujte, výsledky zaokrouhlete na desetiny.
g) Kolik chlapců bylo v 8. ročníku? Je chlapců více než dívek?
h) Kolik procent počtu všech žáků 8. ročníku tvořili chlapci? A kolik procent dívky?
i) Kolik je v tvé třídě chlapců a kolik dívek?
j) Kolik procent počtu žáků tvé třídy jsou chlapci a kolik procent dívky?

Příklad: Na základní škole v Javornici proběhlo **statistické šetření**, kterého se zúčastnilo 624 žáků této školy. Jedna z otázek byla: „Kolik máš celkem sourozenců?“ Zde jsou výsledky:

Počet všech sourozenců	0	1	2	3	4	5
Počet žáků	15	210	244	95	36	24
Počet žáků v procentech	2,40	33,65	39,10	15,22	5,77	3,85

SLOVA POUŽÍVANÁ VE STATISTICE
Dotazovaní žáci ZŠ Javornice ….……………….. **statistický soubor**
Každý z dotazovaných žáků ……….…………… **statistická jednotka**
Počet sourozenců ……………….………………zvolený **znak**
Čísla 0, 1, 2, 3, 4, 5 …………………………… zjištěné **hodnoty znaku**
0 – 5 …………………………………………… **rozsah znaku**
24 žáků má 5 sourozenců …………..………….. 24 je **absolutní četnost**
(frekvence) hodnoty 5

$\frac{\mathbf{24}}{\mathbf{624}}$ = 0,03846… je ta část dotazovaných žáků, kteří mají pět sourozenců
…………………….. 0,03846 je **relativní četnost** hodnoty 5
0,03846 . 100 = 3,846 je počet procent žáků, kteří mají pět sourozenců
…………………….. 3,846 % je **relativní četnost** hodnoty 5

Každé statistické jednotce přiřazujeme **jedinou hodnotu znaku.**
Součet četností se rovná počtu všech jednotek statistického souboru.
Součet relativních četností vyjádřených v procentech je 100 %.
Pokud relativní četnosti zaokrouhlujeme, nemusí nám 100 % vyjít přesně.

Příklad 2 : Proveďte ve své třídě průzkum, jehož cílem je zjistit, kolik z vás se narodilo ve kterém kalendářním měsíci. Výsledky uspořádejte do tabulky, zapište četnosti a vypočítejte relativní četnosti.

8.2 Základní charakteristiky souboru

Aritmetický průměr hodnot (x´) je součet všech hodnot (x_i) vydělený počtem všech statistických jednotek souboru (n).

$$\boxed{x' = \frac{x_1 + x_2 + \ldots + x_n}{n}}$$

Příklad: V osmé třídě je 25 žáků, kteří byli hodnoceni v pololetí z fyziky takto: osmnáct žáků dostalo jedničku, šest žáků dostalo dvojku a jeden žák trojku. Vypočítej průměrnou známku z fyziky v této třídě.
Řešení:

$$x = \frac{\mathbf{18.1 + 6.2 + 1.3}}{\mathbf{25}} = \frac{\mathbf{33}}{\mathbf{25}} = \mathbf{1{,}32}$$

Žáci osmé třídy byli hodnoceni průměrnou známkou 1,32.

Pozor na činěné závěry na základě aritmetického průměru.

Příklad: Máme chlapecký a dívčí kolektiv. V chlapeckém kolektivu má Karel 70.- Kč, Honza 55.- Kč, Zdeněk 45 Kč, Ota 40 Kč, Zbyněk 30.- Kč , Martin 10. Kč, Michal nemá peníze a Tomáš také nemá peníze. V dívčím kolektivu má Věra 40.- Kč, Zuzka 40.- Kč, Jitka 40.- Kč, Jana 30.- Kč, Petra 30.- Kč, Pavla 25.- Kč, Romana 25.- Kč a Nikola 20 Kč.
Vypočítejte kolik korun mají chlapci a dívky.
Řešení : x´(chlapci) = 31,25 Kč y´(dívky) = 31,25 Kč
Průměrně mají všichni stejně, ale finanční situace je v každé skupině jiná. V čem se liší ?

Příklad 3: Vypočítejte aritmetický průměr všech známek (mimo chování) na svém vysvědčení.

Příklad 4: V prodejně obuvi byly v posledním týdnu následující tržby:
v pondělí …………. 33 247 Kč
v úterý ……………. 32 116 Kč
ve středu …………. 29 875 Kč
ve čtvrtek ………… 35 569 Kč
v pátek …………… 47 321 Kč
v sobotu ………….. 25 997 Kč
Vypočítejte aritmetický průměr tržeb čili průměrnou denní tržbu za poslední týden. Výsledek zaokrouhlete na celé koruny.

Příklad: Kontrolní písemná práce z matematiky v roční dopadla takto :

Třída	9.A	9.B	9.C	9.D
průměrná známka	2,21	1,82	2,33	2,11
počet žáků	28	24	32	30

Určete průměrnou známku dětí celého ročníku .

Řešení : Vzhledem k tomu, že četnost žáků každé třídy není stejná, není možné počítat aritmetický známek jako aritmetický průměr aritmetických průměrů, ale musíme zvolit tento postup.

$$x' = \frac{2{,}21.28 + 1{,}82.24 + 2{,}33.32 + 2{,}11.30}{28 + 24 + 32 + 30} = \mathbf{2{,}14}$$

Uvědomte si, co vyjadřuje výraz **2,21.28**.

Příklad 5: Tomáš jel na výlet k babičce 3 hodiny vlakem rychlostí 40 km/hod a pak půl hodiny autobusem průměrnou rychlostí 60 km/hod. Jakou průměrnou rychlostí cestoval ?

Příklad 6: Jirka se vydal na výlet. Nejdříve šel 30 minut průměrnou rychlostí

6 km/hod na autobus, který okamžitě přijel. Autobusem jel 2,5 hodiny průměrnou rychlostí 60 km/hod. Potom čekal 15 minut na vlak, kterým jel hodinu a čtvrt průměrnou rychlostí 50 km/hod. Na určené místo nakonec došel za půl hodiny průměrnou rychlostí 4 km/hod. Jakou průměrnou rychlostí cestoval ?

Hodnotu šetření s nejvyšší četností nazýváme **modus**.

Medián je hodnota, která leží ve středu tabulky uspořádané od nejmenší do nejvyšší hodnoty šetřeného znaku.
Je-li počet jednotek souboru **liché číslo,** je medián sledovaného znaku ta jeho hodnota, která leží „uprostřed“.
Je-li počet jednotek souboru **sudé číslo**, je medián sledovaného znaku aritmetickým průměrem těch jeho dvou hodnot, které jsou „nejblíže středu“.

Příklad: Na naší škole máme volejbalové družstvo dívek a volejbalové družstvo chlapců. Dívky ve volejbalovém družstvu mají výšky v řadě od největší po nejmenší : 194 cm, 192 cm, 175 cm, 175 cm, 175 cm, 174 cm, 172 cm,
171 cm, 171 cm, 170 cm. Určete : a) aritmetický průměr; b) modus; c) medián.

Řešení :

a) $x' = \frac{194 + 192 + 175.3 + 174 + 172 + 171.2 + 170}{10} = 176{,}9$ cm

b) modus je hodnota 175 cm;

c) medián je hodnota 174,5 cm.

Příklad 7 : Tabulka uvádí rozdělení četností výše čtvrtletní odměny pro 42 pracovníků závodu.

Odměny v Kč	3 000	6 000	10 000	15 000	30 000
četnost	5	18	11	7	1

Vypočítejte aritmetický průměr, modus a medián výše čtvrtletní odměny.

Příklad 8 : V tabulce jsou uvedeny počty žáků ve všech třídách základní školy v Javornici.

Třída	1.A	2.A	3.A	4.A	5.A	6.A	7.A	7.B	8.A	9.A
počet žáků	22	24	22	23	26	28	29	27	30	32

Získané údaje zpracujte takto :

a) Statistický soubor tvoří třídy 1.A až 9.A. Roztřiďte je do skupin podle znaku „počet žáků ve třídě“. Sestavte tabulku, v jejímž prvním řádku budou hodnoty znaku „počet žáků ve třídě“ a ve druhém řádku jejich četnosti.
b) Určete modus a medián tohoto znaku.
c) Vypočítej aritmetický průměr počtu žáků ve třídě.

Příklad 9 : V soutěži jednotlivců získali jednotliví závodníci tyto body : 37, 53, 56, 58, 59, 47, 53, 48, 58, 58, 36, 44, 48, 45, 61, 49, 58, 45, 51, 53, 56, 53, 53, 54, 56, 58, 60, 60, 54, 58.
a) Sestavte tabulku, v jejímž prvním sloupci bude znak „počet získaných bodů“ a v druhém sloupci bude příslušná frekvence.
b) Určete aritmetický průměr získaných bodů.
c) Určete modus a medián získaných bodů.

Rozptyl charakterizuje rozložení hodnot ve vzorku vzhledem k aritmetickému průměru. Čím je menší, tím jsou naměřené hodnoty blíže aritmetickému průměru.

Značíme σ (čteme sigma) $\boxed{\sigma = \frac{\sum (x_i - \bar{x})^2}{n}}$

Rozptyl není závislý na aritmetickém průměru. Vzhledem k tomu, že je vyjádřen ve čtvercích měrných jednotek sledovaného znaku, zavádí se pojem **směrodatná odchylka.**

Směrodatná odchylka je definována jako odmocnina rozptylu.
Značíme s $\boxed{s = \sqrt{\sigma}}$

Vzhledem k tomu, že ani směrodatná odchylka nevyjadřuje vztah k aritmetickému průměru, zavádí se pojem **variační koeficient**.

Variační koeficient

Značíme v_k $\boxed{v_k = \frac{s}{\bar{x}} . 100}$

Variační koeficient se udává v procentech.
Variační koeficient, který je větší než 50% , ukazuje na nesourodost statistického souboru a to v takové míře, že použití aritmetického průměru je už stěží oprávněné.

Pro lepší pochopení si zvolíme lehký příklad s třemi čísly, kdy závěry jsou jasné bez našeho počítání, ale při řešení úkolu s více čísly, např. tisícem čísel, již k závěrům potřebujeme naznačené výpočty.

Příklad : V levé skupině máme tři čísla : 7; 8; 9.
V pravé skupině máme čísla 1; 10, 13. Vypočtěte pro obě skupiny základní statistické údaje.

Řešení :	levá skupina	pravá skupina
aritmetický průměr	**8**	**8**

výpočet rozptylu

x_i	x'	$x_i - x'$	$(x_i - x')^2$		x_i	x'	$x_i - x'$	$(x_i - x')^2$
7	8	-1	1		1	8	-7	49
8	8	0	0		10	8	2	4
9	8	1	1		13	8	5	25
			-------					-------
			2					78

dosadíme do vzorce $\sigma = \frac{\sum (x_i - x')^2}{n}$

$\sigma = \frac{2}{3} = \mathbf{0,67}$ $\qquad$ $\sigma = \frac{78}{3} = \mathbf{26}$

dosadíme do vzorce pro směrodatnou odchylku $s = \sqrt{\sigma}$

$s = \sqrt{0,67} = 0,82$ $\qquad$ $s = \sqrt{26} = 5,1$

závěry : většina čísel se odchyluje od aritmetického průměru (8) o méně než jedna v obou směrech, leží tedy mezi čísly 7 a 9.

většina čísel se odchyluje od aritmetického průměru (8) o více než pět v obou směrech, leží tedy mezi 3 a 13.

dosadíme do vzorce pro variační koeficient $v_k = \frac{s}{x'} . 100$

$v_k = \frac{0,82}{8} .100 = \mathbf{10,2\%}$ $\qquad$ $v_k = \frac{5,1}{8} .100 = \mathbf{63,5\%}$

Máme-li soubor čísel a jedno z nich zcela evidentně je nepravděpodobné nebo vzniklo důsledkem nějaké chyby, tak je možné za určitých podmínek toto číslo vyřadit ze statistického zkoumání, protože jiná chyba by mohla ovlivnit výsledky našeho zkoumání.

Příklad 10 : Při měření určitého předmětu byly změřeny tyto údaje v centimetrech : 2,11 2,01 2,09 2,11 2,02 2,03 2,03 2,10 2,05 2,05. Vypočtěte směrodatnou odchylku a variační koeficient tohoto šetření.

Příklad 11 : Ve střelbě z pistole soutěžila tři družstva. Docílili těchto výsledků (číslo udává vzdálenost zásahu od středu terč v milimetrech). Které družstvo mělo největší rozptyl ?
Družstvo A : 10; 8; 9; 7; 8; 10; 9; 7; 3; 10;
Družstvo B : 3; 10; 6; 7; 8; 9; 7; 3; 9; 10;
Družstvo C : 2; 3; 1; 3; 2; 7; 1; 2; 2; 3;

Výsledky statistických šetření se velmi často vyjadřují pomocí **diagramů**.
Každý diagram vyjadřuje vzájemný vztah mezi dvěma i více proměnnými veličinami pomocí přehledných grafických symbolů (číslice, písmena, matematické symboly, schematické obrázky, čáry, obrazce, tělesa, barvy a jejich odstíny).

Druhy diagramů: bodový, spojnicový, hůlkový (úsečkový), sloupcový, kruhový.

Souhrnná cvičení

1) V Janině třídě je 40 % chlapců. jejich průměrná výška je 145 cm. průměrná výška všech dětí v Janině třídě je 142 cm. Ve své třídě je Jana mezi děvčaty nadprůměrně vysoká, ale je menší než je průměrná výška všech dětí v její třídě. Zjistěte, kolik Jana měří, pokud víte, že její výška je v centimetrech celé číslo.

2) Vypočtěte variační koeficient čísel těchto šetření :
a) 4,2 3,9 3,8 4,1 4,0 4,5 4,3;
b) 29,4 29,3 29,5 29,3 29,8;
c) 16,2 15,8 14,8 16,1 16,0 15,9 15,8 15,7 16,1;

Výsledky příkladů

1) a) 1 082 415; **b)** 113 026; **c)** v 2. ročníku, v 9. ročníku; **d)** v 1.- 5. ročníku;
e) čísla ve sloupečku počet žáků; **f)** ano; **g)** chlapců – 58 338, bylo jich více než děvčat;
h) chlapci 51,61 %, dívky 48,39%;
4) 34 021 Kč; **5)** 42,8 km/hod; **6)** 43,5 km/hod; **7)** 8 762.- Kč; 6 000.- Kč; 6 000.- Kč;
8) b) 22, 26,5; **c)** 26,3 **9) b)** 52,63 bodů; **c)** 58 bodů, 53,5 bodů;
10) s = 0,0368781; v_k = 1,79; **11) B : 5,956**; A : 4,09; C : 0,264;

Výsledky souhrnných cvičení

1) 141 cm; **2) a)** 5,36 %; **b)** 0,644%; **c)** 2,27%;