

Možnosti využití statistiky a teorie zpracování dat v práci učitele na 1. stupni ZŠ

Význam statistiky je v soudobé společnosti všeobecně uznáván. Svědčí o tom články v denním i odborném tisku, slýcháme o ní často i ve vystoupeních hospodářských i politických pracovníků.

Definice statistiky je uváděna jako obor zabývající se zkoumáním a kvantitativní charakteristikou hromadných jevů nebo jako číselná evidence hromadných jevů. Existuje několik definic statistiky, které se od sebe často liší. Pro obor statistika jsou příznačné dvě skutečnosti: jejím předmětem jsou hromadné jevy, to jsou jevy, které mají hromadný výskyt a jejím charakteristickým rysem je číselné a grafické vyjadřování zkoumaných skutečností.

Znamená to, že statistika se nezabývá jevy jedinečnými, neopakovatelnými. Zabývá se pouze jevy, které jsou příznačné pro velký počet jedinců, popřípadě, které se často opakují.

Zkoumané skutečnosti o hromadných jevech statistika vyjadřuje statistickými údaji-statistickými daty. Se statistickými údaji se setkáváme např. ve Statistické ročence České republiky.

Slovo statistika, statistický, jsou slova mezinárodní. V převážné většině jazyků je odvozeno od slova stát (např. v němčině: die Statistik-statistika, statistisch-statistický; der Staat-stát; v angličtině: statistics-statistika, statistical-statistický; state-stát).

Statistické informace o sociálním, ekonomickém a ekologickém vývoji České republiky a jejich jednotlivých částí zajišťuje státní statistická služba, kterou vykonává Státní statistický úřad.

1.1. Základní pojmy statistiky

Úkolem statistiky je zkoumání stavu a vývoje hromadných jevů a souvislostí mezi nimi. Jevy rozumíme objekty (předměty), jejich vlastnosti i vzájemné vztahy mezi nimi; pojmem hromadný zdůrazňujeme, že číselné statistické údaje neposkytují informace o individuálních objektech, ale o celých souborech těchto objektů.

1.1.1. Statistický soubor a statistická jednotka

Množinu všech předmětů pozorování (osob, věcí, jevů apod.) shromážděných na základě toho, že mají společné vlastnosti, nazýváme statistickým souborem. Jednotlivé prvky této množiny se nazývají prvky (elementy) statistického souboru nebo též statistické jednotky. Počet všech prvků statistického souboru se nazývá rozsah souboru N .

Soubor, který je předmětem zkoumání, se nazývá základní soubor. Často nelze nebo není účelné provést zkoumání všech statistických jednotek tohoto základního souboru. Základní soubor pak zkoumáme pomocí statistických jednotek, které z něj byly určitým způsobem vybrány a které tvoří tak zv. výběrový soubor.

Příklad 1: Při zjišťování výšky žáků ve třídě je statistickým souborem množina žáků dané třídy. Jejich společnou vlastností je, že jsou žáky například třídy 4. A základní školy ul. Boženy Němcové v Litoměřicích, a že budeme zkoumat jejich výšku. Statistickou jednotkou je žák 4. A základní školy v ul. Boženy Němcové v Litoměřicích. Rozsahem souboru je počet žáků dané třídy, například 20. Statistickým souborem může být také množina všech žáků dané školy.

1.1.2. Statistické znaky

Vlastnosti statistických souborů, které jsou předmětem statistického zkoumání, sleduje statistika prostřednictvím vlastností statistických jednotek daného souboru, které postihuje statistickými znaky.

Statistický znak je vyjádřením určité vlastnosti statistických jednotek (prvků množin) sledovaného statistického souboru; slouží k charakterizování sledovaného hromadného jevu-vlastnosti daného statistického souboru.

Znak (argument) souboru se zpravidla značí x . Jednotlivé údaje znaku se nazývají hodnoty znaku, značí se x_1, x_2, x_N , kde N je rozsah souboru.

Například: Například při určování výšky žáků dané třídy je statistickým znakem výška žáků, hodnotou znaku je číselně vyjádřená příslušná výška žáka např. 142 cm.

Hodnoty znaku mohou být vyjádřeny buď čísly nebo jiným způsobem (zpravidla slovním popisem). V prvním případě mluvíme o znacích kvantitativních, např. tělesná výška, tělesná hmotnost, počet obyvatel měst, atp.. V druhém případě mluvíme o znacích kvalitativních, které se mohou vyskytovat ve dvou druzích (znaky alternativní, např. muž-žena, voják-nevoják, prospěl-neprospěl) nebo ve více druzích (např. povolání, národnost, náboženství, atp.).

Statistickým zkoumáním určitého statistického souboru získáme zpravidla velký počet údajů o jeho statistických znacích. Aby byl tento statistický přehledný je třeba ho někdy nejprve utřídit. Mluvíme o statistickém třídění, které spočívá v rozčlenění statistického souboru do menších skupin, které nazýváme třídami. Je však třeba dodržet zákony třídění, t.j. že třídění musí být úplné - musíme roztrždit všechny hodnoty znaků souboru, musí být disjunktní, průnik skupin musí být prázdná množina a musíme třídit dle určitého znaku.

Třídění podle jednoho znaku může být provedeno v podstatě dvěma způsoby. Při prvním z nich se do jedné třídy shrnují všechny prvky, jimž přísluší táž hodnota znaku x_k

($k = 1, 2, \dots, m$, kde m počet všech různých hodnot znaku x). Hodnota x_k se nazývá třídící znak k -té třídy. Druhý způsob tvoření tříd se používá tehdy, je-li počet různých hodnot znaku příliš velký. Volí se t.j. intervaly.

1.1.3. Četnost

Počet prvků statistického souboru, které patří do k -té třídy se nazývá četnost (absolutní četnost) prvků v k -té třídě; značíme N_k . Podíl $r_k = \frac{N_k}{N}$, kde

N je rozsah uvažovaného statistického souboru, se nazývá relativní (poměrná četnost) prvků v k -té třídě. Vyjadřuje se jako číslo desetinné nebo v procentech. Součet všech četností prvků v první až k -té třídě se nazývá kumulativní četnost prvků v k -té třídě. Součet všech relativních četností prvků v první až k -té třídě se nazývá kumulativní relativní četnost prvků v k -té třídě.

Příklad 2: Ve 4. A třídě základní školy ul. B. Němcové v Litoměřicích byly naměřeny tyto výšky žáků: 130 cm, 132 cm, 135 cm, 135 cm, 138 cm, 140 cm, 142 cm, 142 cm, 142 cm, 142 cm, 147 cm, 147 cm, 149 cm, 149 cm, 152 cm, 159 cm, 160 cm, 160 cm, 164 cm, 164 cm.

Hodnoty sledovaného znaku t.j. výšku žáků třídy 4. A základní školy ul.

B.Němcové v Litoměřicích uvedeme v tabulce rozdělení četností neboli četnostní tabulce:

Hodnoty znaku x_i	Četnosti		Kumulativní četnosti	
	absolutní	relativní	absolutní	relativní
130	1	0,05	1	0,05
132	1	0,05	2	0,10
135	2	0,10	4	0,20
138	1	0,05	5	0,25
140	1	0,05	6	0,30
142	4	0,20	10	0,50
147	2	0,10	12	0,60
149	2	0,10	14	0,70
152	1	0,05	15	0,75
159	1	0,05	16	0,80
160	2	0,10	18	0,90
164	2	0,10	20	1,00
Součet	20	1,00		

1.2. Charakteristiky statistického souboru

Statistickými charakteristikami nazýváme čísla, která podávají stručnou základní informaci o uvažovaném statistickém souboru z různých hledisek. Je-li předmětem našeho zájmu jediný kvantitativní znak, jde o charakteristiku úrovně (polohy) a charakteristiku variability (proměnnosti, rozptýlení).

1.2.1. Charakteristiky úrovně (polohy)

Charakteristiky úrovně (polohy) jsou čísla, která charakterizují „průměrnou hodnotu“ sledovaného kvantitativního znaku. Patří mezi ně aritmetický průměr, medián, modus, případně harmonický průměr a geometrický průměr.

1.2.1.1. Aritmetický průměr

Aritmetický průměr $\bar{x}$ hodnot $x_1, x_2, \dots, x_N$ kvantitativního znaku x je součet těchto hodnot dělený jejich počtem (rozsahem souboru) N :

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

Příklad 3: Jsou dány výšky žáků (viz příklad ze str.2). Určete jejich aritmetický průměr:

$$\bar{x} = \frac{130+132+135+135+138+140+142+142+142+142+147+147+149+149}{20} + \frac{152+159+160+160+164+164}{20} = \frac{2929}{20} = 146,45$$

Aritmetický průměr výšky žáků je 146,45 cm .

Máme-li sestavenou tabulku rozdělení četností, podle níž hodnota x_1 má četnost N_1 , hodnota x_2 má četnost N_2 , ... ,hodnota x_m má četnost N_m (kde $N_1 + N_2 + \dots + N_m = N$), vypočteme aritmetický průměr podle vzorce:

$$\bar{x} = \frac{N_1 \cdot x_1 + N_2 \cdot x_2 + \dots + N_m \cdot x_m}{N} = \frac{1}{N} \sum_{k=1}^m N_k \cdot x_k$$

Říká se mu vážený aritmetický průměr (jednotlivé hodnoty znaku jsou „váženy“ jejich četnostmi).

Příklad 4: Určete vážený průměr předchozího příkladu:

$$\bar{x} = \frac{1 \cdot 130 + 1 \cdot 132 + 2 \cdot 135 + 1 \cdot 138 + 1 \cdot 140 + 4 \cdot 142 + 2 \cdot 147 + 2 \cdot 149 + 1 \cdot 152 + 1 \cdot 159 + 2 \cdot 160 + 2 \cdot 164}{20} = 146,45$$

Vážený průměr výšky žáků je 146,45 cm .

Aritmetický průměr má tyto důležité vlastnosti:

- Součet rozdílů (tzv. odchylek) $x_i - \bar{x}$ jednotlivých hodnot znaku x_i od jejich aritmetického průměru $\bar{x}$ se rovná nule.
- Přičteme-li nebo odečteme-li od každé hodnoty znaku konstantní číslo A, pak aritmetický průměr nových hodnot se rovná aritmetickému průměru původních hodnot zvětšenému nebo zmenšenému o číslo A.
- Násobíme-li nebo dělíme-li každou hodnotu znaku určitou konstantou K různou od nuly pak aritmetický průměr nových hodnot se rovná aritmetickému průměru původních hodnot násobenému nebo dělenému konstantou K.

3.2.1.2. Medián ($\hat{x}$)

Medián u souborů, jejichž rozsah N je liché číslo, je roven hodnotě znaku prostředního prvku [tj. $\left(\frac{N+1}{2}\right)$ - tého prvku] a u souborů, jejichž rozsah N je sudé číslo, je roven aritmetickému průměru hodnot dvou středních prvků [tj. $\frac{N}{2}$ - tého a $\left(\frac{N}{2}+1\right)$ - tého prvku] .

Poznámka: Hodnoty znaků musí být seřazeny podle velikosti.

Příklad 5 : Medián výšek žáků 4.A je $\hat{x} = \frac{142+147}{2} = 144,5$ centimetrů.

1.2.1.3. Modus ($\tilde{x}$)

Modus je hodnota znaku, která má maximální četnost.

Příklad 6: Modus výšek žáků 4.A je $\tilde{x} = 142$ cm.

Poznámka: Medián a modus se používají jako charakteristiky úrovně, jsou-li extrémní hodnoty znaku mimořádně odlišné od ostatních hodnot znaků, takže aritmetický průměr je netypickou charakteristikou úrovně souboru.

1.2.1.4. Harmonický průměr

Harmonický průměr $\bar{x}_H$ hodnot znaků $x_1, x_2, \dots, x_N$ je

$$\bar{x}_H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N}} = N : \sum_{i=1}^N \frac{1}{x_i}$$

Máme-li sestavenou tabulku rozdělení četností, podle níž hodnota x_1 má četnost N_1 , hodnota x_2 má četnost N_2 , ..., hodnota x_m má četnost N_m (kde $N_1 + N_2 + \dots + N_m = N$), vypočteme vážený harmonický průměr podle vzorce:

$$\bar{x}_H = \frac{N_1 + N_2 + \dots + N_m}{N_1 \cdot \frac{1}{x_1} + N_2 \cdot \frac{1}{x_2} + \dots + N_m \cdot \frac{1}{x_m}} = N : \sum_{k=1}^m \frac{N_k}{x_k}$$

1.2.1.5. Geometrický průměr

Geometrický průměr $\bar{x}_G$ hodnot znaků $x_1, x_2, \dots, x_N$ je

$$\bar{x}_G = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$$

3.2.2. Charakteristiky variability (proměnnosti, rozptýlení)

Charakteristiky variability (proměnnosti, rozptýlení) hodnot znaků jsou čísla, která charakterizují z různých hledisek proměnnost sledovaného kvantitativního znaku statistického souboru. Patří mezi ně variační rozpětí, průměrná odchylka, rozptyl, směrodatná odchylka a variační koeficient.

3.2.2.1. Variační rozpětí

Variační rozpětí R je rozdíl mezi největší a nejmenší hodnotou znaku prvků daného souboru:

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

Příklad 7 : Variační rozpětí výšek žáků 4. A je $R = 164 - 130 = 34$ cm.

1.2.2.2. Průměrná odchylka

Průměrná odchylka $\bar{d}$ je aritmetický průměr absolutních hodnot odchylek znaků všech prvků souboru od jejich aritmetického průměru :

$$\bar{d} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_N - \bar{x}|}{N}$$

1.2.2.3. Relativní průměrná odchylka

Relativní průměrná odchylka r je podíl průměrné odchylky a příslušného aritmetického průměru:

$$r = \frac{\bar{d}}{\bar{x}}$$

1.2.2.4. Rozptyl

Rozptyl s^2 je aritmetický průměr druhých mocnin odchylek hodnot znaku od aritmetického průměru:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

3.2.2.5. Směrodatná odchylka

Směrodatná odchylka s je druhá odmocnina z rozptylu.

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

1.2.2.6. Variační koeficient

Variační koeficient V je poměr směrodatné odchylky a aritmetického průměru vyjádřený v procentech:

$$V = \frac{s}{\bar{x}} \cdot 100$$

Příklad 8 :

Je dán statistický soubor 5, 4, 5, 5, 8, 3, 6, 5, 5, 7, 9 . Určete charakteristiky úrovně (aritmetický, geometrický a harmonický průměr, modus a medián) a charakteristiky variability (variační rozpětí, průměrnou odchylku, rozptyl, směrodatnou odchylku a variační koeficient daného souboru.

Řešení:

Seřaďme nejprve hodnoty znaku podle velikosti (viz. poznámka u výpočtu mediánu).

Aritmetický průměr

$$\bar{x} = \frac{3+4+5+5+5+5+5+6+7+8+9}{11} = \frac{62}{11} = 5,636 \text{ .}$$

Geometrický průměr

$$\bar{x}_G = \sqrt[11]{3 \cdot 4 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9} = 5,398$$

Harmonický průměr

$$\bar{x}_H = \frac{11}{\frac{1}{3} + \frac{1}{4} + 5 \cdot \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9}} = 5,167$$

Modus $\tilde{x} = 5$

Medián $\hat{x} = 5$

Variační rozpětí

$$R = 9 - 3 = 6$$

Průměrná odchylka

$$\bar{d} = \frac{|5,636 - 3| + |5,636 - 4| + 5 \cdot |5,636 - 5| + |6 - 5,636| + |7 - 5,636| + |8 - 5,636| + |9 - 5,636|}{11} = 1,355$$

Rozptyl

$$s^2 = \frac{2,636^2 + 1,636^2 + 5 \cdot 0,636^2 + 0,364^2 + 1,364^2 + 2,364^2 + 3,364^2}{11} = 2,777$$

Směrodatná odchylka

$$s = \sqrt{2,777} = 1,666$$

Variační koeficient

$$V = \frac{1,666}{5,636} \cdot 100 = 29,599$$

1.3. Teorie statistické závislosti. Dvourozměrný statistický prostor.

Dvourozměrným statistickým souborem se nazývá statistický soubor, u něhož je předmětem zájmu vztah dvou statistických znaků. Například je to vztah klasifikace z matematiky v pololetí a klasifikace kontrolní písemné práce z matematiky ve 2. pololetí v dané třídě.

Nejjednodušší závislostí mezi těmito dvěma statistickými znaky je lineární závislost, jejíž těsnost se měří korelačním koeficientem r_{xy} , který nabývá hodnot v intervalu od -1 do 1.

Korelační koeficient rovný 1 značí přímou úplnou funkční lineární závislost, kdežto rovná-li se -1, jde o nepřímou úplnou funkční lineární závislost. Nabude-li korelační koeficient hodnoty 0, značí to, že lineární závislost není pro aproximaci vhodná. Při hodnotách -1 až 0 nebo 0 až 1 jde o větší či menší stupeň těsnosti nepřímé či přímé lineární závislosti.

Vzorec pro korelační koeficient:

kde s_{xy} se nazývá kovari-
datné odchyly znaku x a

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

ance , s_x a s_y jsou směro-
znaku y .

Vzorec pro kovarianci je

$$s_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N}$$

Vzorec pro korelační koeficient r_{xy} tedy je

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N \sqrt{\left[\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right] \cdot \left[\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \right]}}$$

Příklad 9:

Závěrečného testu z matematiky se zúčastnilo 12 žáků čtvrtého ročníku. Výsledky testu (vyjádřené průměrnou známkou) i jejich pololetní klasifikaci z matematiky vyjadřuje následující tabulka. Vypočítejte korelační koeficient mezi pololetní klasifikací a průměrnou známkou z testu.

Žák	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
pol. klasifikace	2	1	3	2	1	4	1	2	2	3	3	3
průměrná známka	1	1,16	2,17	1,7	1,16	1,83	1,33	1,16	1,16	1,16	1,5	1,83

Řešení:

Označíme-li pololetní klasifikaci x a průměrnou známkou z testu y pak $\bar{x} = 2,25$ a $\bar{y} = 1,43$.

$$s_{xy} = \frac{(-0,25) \cdot (-0,43) + (-1,25) \cdot (-0,26) + (0,75) \cdot (0,74) + (-0,25) \cdot (0,27) + (-1,25) \cdot (-0,27) + (1,75) \cdot (0,4)}{12} +$$

$$+ \frac{(-1,25) \cdot (-0,1) + (-0,25) \cdot (-0,27) + (-0,25) \cdot (-0,27) + (0,75) \cdot (-0,27) + (0,75) \cdot (0,07) + (0,75) \cdot (0,4)}{12} = 0,1973$$

$$s_x = \sqrt{\frac{1}{12} [3 \cdot (1 - 2,25)^2 + 4 \cdot (2 - 2,25)^2 + 4 \cdot (3 - 2,25)^2 + 1 \cdot (4 - 2,25)^2]} = 0,9242$$

obdobně $s_y = 0,3541$

a korelační koeficient

$$r_{xy} = \frac{0,1973}{0,9242 \cdot 0,3541} = 0,6029$$

Vypočtená hodnota korelačního koeficientu potvrzuje přímou lineární závislost mezi pololetní klasifikací a hodnocením testu.

Shrňme možnosti využití statistiky:

- 1. Vlastní výuka žáků s využitím statistiky (statistický soubor vytváří sami žáci)**
- 2. Využití statistiky ve vlastní práci učitele (klasifikace, docházka do školy, atp., statistický soubor je dán)**

Poznámka:

Statistický soubor vytvořený z výšek dětí nebo jejich hmotnosti není z výchovného hlediska vhodný. Děti se pak třeba tomu kdo má největší hmotnost posmívají nebo jsou-li nejmenší mají komplex méněcennosti. Student A vymyslel graf úsměvů. Do třídy přinesl provázek, dětem nastříhal příslušné kousky tohoto provázku. Děti si přinesly nůžky a každý si změřil svůj úsměv. Usmály se a provázek vedly od jednoho koutku úst přes zuby k druhému koutku a zde provázek ustříhly. Pan učitel měl připravený seznam dětí a lepidlo a každé dítě ke svému jménu nalepilo svůj provázek jako model úsečky. Počátek byl určen svislou čarou. Nyní děti přišly samy na to, že bude dobře provázky změřit na milimetry. Postupovalo se oběma směry, nejdřív graf a pak číselná data.

Graf úsměvů

Jméno	Úsměv	Délka v mm
Adam		48
Bedřich		53
David		50
Emil		45
	A tak dále	

Pan učitel vyzval děti, aby vymýšlely vlastní otázky. Děti kladly otázky. Kdo má ze třídy nejdelší úsměv? Kdo má nejkratší úsměv. Jaký je rozdíl mezi nejdelším a nejkratším úsměvem. Jak dlouhý je úsměv celé třídy? Jak dlouhý je úsměv chlapců ve třídě? Jak dlouhý je úsměv děvčat ve třídě? Zde uvedly, že je to „nespravedlivé“, neboť děvčat je ve třídě méně. Tak tedy rozhodly, že spočítají průměrný úsměv na jednoho chlapce a na jednu dívku. Zde přišly samy na aritmetický průměr. Jaké mají děti odchylky od aritmetického průměru? A tak podobně. Matematika byla jakousi matematickou laboratoří. Charakteristiky statistického souboru dětí vytvářely vlastní tvůrčí práci, šlo vlastně o didaktický konstruktivismus.