

Základy popisné statistiky

Studijní materiál do didaktiky matematiky

Růžena Blažková

Rámcový vzdělávací program pro základní vzdělávání obsahuje v části Závislosti, vztahy, práce s daty očekávané výstupy ve 2. období 1. stupně ZŠ:

- Žák vyhledává, sbírá a třídí data
- Žák čte a sestavuje jednoduché tabulky a diagramy.

1. Úvod

Se statistikou se každý člověk během svého života někdy setká. Může využívat jejích **výsledků**, např. číselných informací o hromadných společenských jevech, např. z tisku nebo dalších médií. Nebo může být **objektem** šetření, např. při sčítání lidu, statistických šetření různých agentur, různých anket. Nebo využívá **výsledků** statistiky (např. průměrný prospěch žáka, třídy, školy). Také může využívat **metod** statistiky, kdy zpracovává statistické údaje při své profesní činnosti (např. vyhodnocení výzkumného šetření).

Statistika má význam z hlediska společenského, neboť umožňuje poznávat dosavadní vývoj společenských jevů a na základě tohoto poznání umožňuje činit určitá rozhodnutí. Typickým příkladem je současná doba koronavirová.

Statistika je soubor metod, které nám umožňují činit rozumná rozhodnutí v případě nejistoty. Je založená na pozorování, porovnávání, posuzování a zhodnocení množství informací. Moderní statistika tvoří základ teorie rozhodování.

2. Historická poznámka

Původ slova „statistika“ pochází z latiny (status – stát) a nejprve představovala nauku o státu. Blíže k dnešnímu pojetí statistiky měla tzv. anglická politická aritmetika (zakl. J. Graunt a W. Petty), která se zabývala shromažďováním číselných údajů o ekonomických a demografických jevech. Počátky moderní statistiky jsou kladeny do 19. století a jsou spojovány se jménem Belgičana Adolfa Quételeta, který se zabýval číselně vyjádřitelnými vlastnostmi společnosti. Další význam pro rozvoj statistiky mělo založení anglické statistické školy (aplikace v biologii, zemědělství – F. Galton, K. Pearson, R. A. Fischer. Na vývoji metod matematické statistiky mají od počátku 20. století významný podíl B. Gosset (pseudonym Student), P. Čebyšev, A. Ljapunov, A. Markov, Kolmogorov, Bernstejn, Romanovskij a další.

Ve vývoji statistika nastala významná proměna ve 30. letech, kdy vzniká moderní, analytická, induktivní statistika, jejímž základním pojmem je **výběr**. S použitím matematických metod se stala samostatným vědním oborem.

3. Základní pojmy

Statistika je vědní obor, který se zabývá hromadným zkoumáním, pozorováním či šetřením určitých objektů a jevů. Zkoumá jevy společenské, přírodní, technické.

Statistická šetření: - sestavování statistických výkazů hlášení či soupisů
- organizované šetření (ankety, výzkumy veřejného mínění, sčítání lidu aj.)
- studování jevu (měření předmětů, sledování teploty, ceny výrobků aj.).

Statistické šetření se provádí na určité množině – dostatečně rozsáhlé skupině případů, na statistickém souboru.

Statistický soubor je množina – skupina prvků (objektů, osob, událostí aj.), které mají určité společné vlastnosti.

Rozlišujeme statistické soubory **základní a výběrové**. Základním statistickým souborem může být např. všechno obyvatelstvo světa, všechno vodstvo na Zemi apod. Vymezení základního souboru může někdy přinášet problémy, šetření na celém souboru může být časově náročné nebo i nemožné, proto v praxi používáme soubor výběrový (podmnožina statistického souboru). Tento soubor by měl vypovídat o základním souboru, z kterého byl odvozen (jinak dochází ke zkreslení výsledků).

Statistický soubor lze rozdělit na dvě části: Na část, ve které nastává zkoumaný jev a část, ve které zkoumaný jev nenastává. Základním statistickým úkonem je třídění, které provádíme podle jistých kritérií (rozklad množiny na třídy). Respektujeme zásadu úplnosti (každý prvek statistického souboru musí být v některé třídě), a zásadu jednoznačnosti (žádný prvek nesmí být současně ve dvou třídách). Třídění může být dichotomické, trichotomické, obecně multitonické (např. hledání v klíči pro určování rostlin).

Prvky statistického souboru se nazývají **statistické jednotky**.

Počet jednotek statistického souboru se nazývá **rozsah souboru**

Každá statistická jednotka je nositelem určitých vlastností. Ty vlastnosti, které jsou důležité z hlediska účelu provádění určitého statistického zkoumání, se nazývají **statistické znaky**. Statistické jednotky tedy vyšetřujeme z hlediska určitého znaku nebo několika znaků, které si zvolíme.

Statistické znaky dělíme na **kvantitativní** (číselné) a **kvalitativní** (slovní).

Některé kvantitativní znaky mohou nabývat pouze jednotlivých izolovaných hodnot - **diskrétní** znaky (např. počet obyvatel obce), nebo nabývají libovolných reálných hodnot z určitého intervalu – **spojité** znaky (např. hektarové výnosy).

V případě, že kvantitativní znak nabývá pouze dvou variant, hovoříme o znaku **alternativním** (např. muž, žena), nabývá-li více variant, hovoříme o znaku **multiplikativním** (např. kvalifikace, státní příslušnost).

Číslo, které udává, kolikrát se daná hodnota znaku ve statistickém souboru vyskytuje, se nazývá **absolutní četnost** hodnoty znaku. Součet jednotlivých četností sledovaného znaku je roven rozsahu souboru.

$$n_1 + n_2 + \dots + n_k = n$$

Poměrná – **relativní četnost** jevu je poměr absolutní četnosti a rozsahu souboru. Zpravidla se označuje řeckým písmenem „ ν “ - „ ν “.

$$\nu_k = \frac{n_k}{n}$$

Součet relativních četností je roven jedné.

Relativní četnosti lze vyjadřovat také v procentech, pak je jejich součet 100%.

Příklady možných statistických šetření na ZŠ:

| Statistický soubor | Statistická jednotka | Statistický znak |
|---|--------------------------------|--|
| Všichni žáci třídy | Žák třídy | Výška žáka Hmotnost žáka Prospěch v matematice Záliby |
| Všichni žáci školy | Žák školy | Studium jazyků Zařazení do sportovních aktivit |
| Všechny dopravní prostředky, které projedou kolem určitého stanoviště | Jednotlivý dopravní prostředek | Druh vozidla Typ vozidla Barva Poznávací značka |
| Všechny hody hrací kostkou | Jednotlivý hod | Počet ok na jedné stěně |
| Všechna slova na jedné straně knihy | Jedno slovo | Počet písmen Slovní druh |

| | | |
|--|--------------|---|
| | | |
| Všechny dopravní nehody v jednom roce v ČR | Jedna nehoda | Příčiny nehod Hmotná škoda Zranění osob |

4. Diagramy

Rozdělení četností znaků vyjadřujeme buď v tabulce nebo graficky pomocí diagramů.

Diagram vyjadřuje vzájemný vztah mezi dvěma či více proměnnými veličinami pomocí přehledných grafických symbolů. Rychle a názorně poskytne obrazovou informaci o studovaném jevu.

Diagram **obrázkový** – obrázek vyjadřuje určitý počet prvků, např. obrázek jednoho auta představuje např. 1 000 vyrobených aut.

Diagram **bodový**- četnosti jsou znázorněny pomocí izolovaných bodů.

Diagram **sloupkový** – **histogram** – používá se v případech, kdy jsou hodnoty znaků sdruženy do intervalů. Tyto intervaly tvoří jednu stranu sloupků (obdélníků), druhou stranu tvoří četnosti.

Diagram **hůlkový** – **úsečkový** – četnosti znaků jsou znázorněny úsečkami

Diagram **spojicový** – **polygon četností** – získá se spojením bodů, jejichž souřadnice tvoří hodnota kvantitativního znaku a odpovídající četnost.

Diagram **kruhový** – různým hodnotám znaků odpovídají kruhové výseče. Jednomu procentu relativní četnosti odpovídá středový úhel $3,6^\circ$.

Ve sdělovacích prostředcích (televize) se využívá prostorového znázornění statistických údajů (kvádry, válce apod.).

Ke znázorňování diagramů je výhodné využít Excel.

5. Charakteristiky polohy

Aritmetický průměr

Aritmetický průměr je definován jako podíl součtu hodnot znaku zjištěných u všech jednotek souboru a počtu všech jednotek souboru:

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

Vlastnosti aritmetického průměru:

1. Matematické vyjádření aritmetického průměru je jednoduché a snadno použitelné pro odvození dalších vztahů.
2. Výpočet je založen na všech pozorovaných hodnotách.
3. Součet všech odchylek jednotlivých hodnot od aritmetického průměru je vždy roven nule.
4. Aritmetický průměr je ovlivňován krajními hodnotami.

Příklad – výpočet průměrné mzdy. Pokud máme 5 pracovníků a jejich mzdy nejsou příliš rozptýleny, aritmetický průměr je seriózní informací:

$$x_a = \frac{12000 + 13000 + 14000 + 15000 + 16000}{5} = 14000$$

Jestliže např. jen jeden má výrazně větší mzdu než ostatní, aritmetický průměr nevyovídá seriózně o souboru:

$$x_a = \frac{12000 + 13000 + 14000 + 150000 + 16000}{5} = 41000$$

Modus

Modus znaku x je hodnota s největší četností, značí se $Mod(x)$. Udává, který výsledek je zastoupen nejvíce, nepodává informace o krajních hodnotách. Praktický význam má např. pro oděvní a obuvnický průmysl (které velikosti se v populaci vyskytují nejvíce).

Medián

Medián je prostřední hodnota znaku, jsou-li hodnoty uspořádány podle velikosti. Značí se $Med(x)$. Je to nejrychleji zjistitelná střední hodnota má před sebou i za sebou stejný počet hodnot. U lichého počtu hodnot je to prostřední hodnota, u sudého počtu je to aritmetický průměr prostředních dvou.

Pro úplnost uvádíme názvy dalších průměrů, se kterými se na 1. stupni ZŠ nesetkáme.

Harmonický průměr

Geometrický průměr

Vážený průměr

Kvadratický průměr

Můžeme se setkat i s dalšími pojmy, jako jsou rozptyl, směrodatná odchylka.

5. Metody práce

Konkrétní statistické šetření ve třídě, na něm se vysvětlí potřebné pojmy

Práce se statistickou ročenkou (každoročně vydávaná Českým statistickým úřadem)

Práce s kalkulátory

Práce se statistickými počítačovými programy

Údaje Českého statistického úřadu (internet, denní tisk)

Práce s počítačem, např. Excel

Didaktický aspekt

Na konkrétních příkladech ilustrovat co nejlépe význam, cíle a možnosti matematické statistiky.

Volíme činnosti, na kterých se podílejí všichni žáci třídy, vybíráme témata, která jsou pro dané žáky atraktivní (aktivizace žáků na volbě statistických šetření i jejich zpracování).

Využíváme možnosti projektové výuky.

Statistické šetření provádíme ve třech etapách:

- Pozorování, šetření, měření, provádění pokusů, zjišťování údajů, sběr dat
- Zpracování údajů
- Rozbor výsledků, závěry, návrhy.