Pedagogická fakulta
Faculty of Education
Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Pedagogická fakulta

Katedra matematiky

Bakalářská práce

Základy statistického zpracování dat

Vypracoval: Miroslav Janošťák

Vedoucí práce: doc. RNDr.**Vladimíra Petrášková**,Ph.D.

České Budějovice 2020

Prohlášení

Prohlašuji, že svoji bakalářskou práci na téma Statistika v příkladech pro vzdělávání na středních školách jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě, elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích …………. ……………

Poděkování

Rád bych touto cestou poděkoval doc. RNDr**. Vladimíře Petráškové,** Ph.D. za odborné vedení a podnětné rady.

Anotace

Cílem práce je vytvořit sbírku řešených příkladů ze statistiky, která bude určena studentům středních škol. Příklady budou ukazovat jednoduché užití statistiky v praxi. Práce bude zaměřena na následující témata: statistický soubor, jednotka, znak (nominální, ordinální, metrické), rozložení četností (absolutní, relativní, kumulativní absolutní, kumulativní relativní) a jeho grafické znázornění (polygon, histogram, ogiva), dále aritmetický, harmonický a geometrický průměr, modus, medián, kvartily, směrodatná odchylka, relativní odchylka, variační koeficient a kvartilová odchylka.

Abstract

The aim of this thesis is to create didactic material of basic statistics which is intented for high school students. The work will be focused on the following topic: statistical file, unit, symbol, frequency distribution, graphical representation, arithmetic mean, geometric diameter, modus, median, quartiles, etc.

Obsah

Úvod

Cílem mé bakalářské práce je přiblížit základy zpracování statistických dat žákům střední školy pomocí přehledného učebního materiálu s ukázkovými řešenými příklady. Využití způsobů zpracování dat na vzorových příkladech z běžného života. Na jednotlivých příkladech objasním, co znamená četnost, aritmetický průměr, geometrický průměr, harmonický průměr, modus, medián, odchylka směrodatná, relativní nebo kvartilová. Jak se v praxi využívají tyto postupy a k čemu nám slouží.

Toto téma jsem si zvolil, jelikož chybí publikace, kde by byly tyto postupy sjednoceny a vysvětleny na praktických příkladech. Je potřeba vytvářet přehledné materiály pro usnadnění pochopení všech matematických témat. Mým cílem je vytvořit přehledný materiál pro výuku, či samostudium základů statistiky.

1. Statistika

Statistika je věda, která se zabývá sběrem, tříděním a zpracováním dat. Vychází ze shromážděných statistických údajů. Slovo *statistika* pochází z latinského „status“, což znamená *stav*. Se statistickými údaji se setkáváme v běžném životě každý den. V novinách se setkáváme s daty zpracovanými do tabulek i grafů. Statisticky jsou zpracovány sportovní výsledky, nehody, dlouhodobý stav počasí, výsledky ve škole, porodnost i úmrtnost. Každý člověk je zahrnut do statistického souboru při sčítání lidu (Zwerenz 2015).

1.1 Základní pojmy

Stručně si představíme pár základních pojmů, se kterými se budeme ve statistice potkávat a podrobněji si je rozebereme v následujících kapitolách.

Statistická data – jsou údaje o tzv. statistických jevech, např. společenských, přírodních, technických aj. Statistická data jsou informace o jednom prvku statistického souboru, lze je rozdělit do několika skupin podle jejich charakteru.

Statistický soubor – je soubor osob, věcí, událostí, časových období, jevů apod. shromážděných na základě určité společné vlastnosti. Dílčí prvky se nazývají jednotky.

Statistické jednotky – jsou jednotlivé prvky statistického souboru.

Rozsah souboru – je počet všech prvků statistického souboru (Calda a Dupač 1993).

1.2 Data

Nasbírané informace označujeme jako data. Obvykle se jedná o číselné údaje. Data se musí dále před samostatnou prezentací rozdělit do skupin neboli tříd, aby seznamy údajů dávaly nějaký smysl. Data dále prezentujeme přehledným způsobem v podobě tabulek nebo grafů. Nezpracovaným nasbíraným informacím říkáme syrová data.

Všechna data je třeba nasbírat a roztřídit před samotnou prezentací a analýzou dat. Nejběžnější způsob sběru dat je průzkum. Určité skupině lidí je položena otázka, dotazník týkající se jejich názorů, návyků či preferencí právě v podobě dotazníku. Jejich odpovědi představují syrová data, která pak třídíme do tabulek a znázorňujeme pomocí grafů.

Dotazník bývá sestaven nejčastěji formou více výběrového testu, kdy si vybíráte odpověď z předem daných možností. Odpovědi se poté snadněji rozdělují do jednotlivých tříd dat (Vordermanová 2015).

Data můžeme dělit dle typu měření na nominální, ordinální a metrická.

1.2.1 Nominální data

Nominální data mají význam jisté kvality, proto někdy mluvíme o kvalitativních datech. Nedají se seřadit, nemají žádnou velikost. Obvykle je na výběr pouze z konečné množiny možností. Typickým příkladem je pohlaví, rodinný stav nebo krevní skupina. Jelikož možnosti rozdělujeme do určitých kategorií, můžeme hovořit o datech kategoriálních. Jednotlivé hodnoty nelze porovnat. Nerozlišujeme jejich velikost. Zvláštní skupinou jsou alternativní data, která nabývají pouze hodnoty Ano a Ne. Lze u nich vypočítat modus. (Podrobněji viz kapitola 3.4 Modus)

V některých případech se používá kódování jednotlivých hodnot čísly, tato čísla je nutné vnímat pouze jako symboly.

1.2.2 Ordinální data

Ordinální data představují podobně jako nominální výběr z určitého počtu možností. Také nenabývají číselných hodnot, rozdíl vytváří možnost data seřadit. U každé vybrané dvojice lze jednoznačně určit, která hodnota je větší a která menší. Příkladem je spokojenost pacienta s lékařskou péčí (spokojen, mírně spokojen, mírně nespokojen, nespokojen). Lze u nich vypočítat modus a medián. (Podrobněji viz kapitola 3.5 Medián)

1.2.3 Metrická data

Metrická data neboli kardinální dělíme na intervalová a poměrová. Intervalová data jsou taková data, ve kterých má smysl hodnotit i vzdálenosti mezi jednotlivými kategoriemi nebo hodnotami. Příkladem může být například teplota. Rozdíl teploty 10 °C je vždy stejný bez ohledu na konkrétní výchozí teplotu. Druhým typem jsou data poměrová. U těchto dat lze určit kolikrát je jedna hodnota menší či větší než druhá. Patří sem například věk, cena výrobku, roční příjem aj. (IASTAT 2001).

2. Statistické veličiny

2.1 Statistický soubor, jednotka a znak

Pro statistiku je charakteristické, že zkoumá společenské, přírodní, či technické jevy na dostatečně rozsáhlém souboru a hledá vlastnosti jevů, které se projevují v celém souboru případů, nikoli na jednom případě. Proto vycházíme ze základního pojmu statistický soubor – osob, věcí, událostí. Jednotlivé prvky statistického souboru se nazývají statistické jednotky. Tyto jednotky analyzujeme z hlediska zvoleného znaku. U každé jednotky zaznamenáváme hodnotu znaku. Hodnoty musí být nastaveny tak, aby byly navzájem neslučitelné a jedna z nich musela vždy nastat. Znaky, které lze rozlišit číselnou velikostí nazýváme kvantitativním znakem. Například výška postavy v cm, měsíční příjem, hmotnost, známky. Je-li něco popsáno číslem, ještě neznamená, že se jedná o kvantitativní znak. Například autobusy jsou očíslovány, ale jedná se pouze o číselné označení linek.

Znaky, jehož hodnoty se liší kvalitou, nazýváme kvalitativními znaky. Například povolání, či druh nemoci. Nejjednodušším příkladem kvalitativního znaku jsou jevy s jevem opačným. Prospěl – neprospěl, muž – žena. Tyto znaky se nazývají alternativní. (Calda a Dupač 1993)

Příklad

Ředitel chce zjistit počet zbývajících dní dovolené u svých 150 zaměstnanců. Urči

a) co je statistický soubor, statistická jednotka, jaký je rozsah souboru

b) společný znak, hodnotu znaku.

Řešení

a)

- Statistickým souborem je množina všech zaměstnanců. Jejich společnou vlastností je, že jsou zaměstnanci firmy.

- Statistickou jednotkou je každý zaměstnanec
- Rozsah souboru je počet zaměstnanců, tedy $n = 150$.

b)

- Statistický znak je „počet zbývajících dnů dovolené“
- Hodnota znaku je číselný údaj, počet dnů dovolené

2.2 Rozložení četností

Při statistickém šetření se zpravidla vyšetřuje řada znaků, které nás zajímají jak každý zvlášť, tak i ve vzájemném vztahu. Ve většině šetření je počet různých hodnot sledovaného znaku menší než počet jednotek tohoto statistického souboru tzn., že několik různých statistických jednotek stejného souboru nabývá stejných hodnot znaku. Například dva lidé mohou mít stejný věk nebo stejnou váhu. Výsledkem šetření je seznam jednotek s udáním hodnoty znaku u každé z nich. Jsou-li jednotky v seznamu očíslovány *1, 2, 3,…., n*, pak jim odpovídající hodnoty znaku x označíme $x_1, x_2, ..., x_n$. Jestliže znak x nabývá pouze určitého počtu r různých hodnot, označujeme je $x_1^*, x_2^*, ..., x_r^*$. Pro každou možnou hodnotu x_j^* pak zjistíme, kolikrát se vyskytla mezi $x_1, x_2, x_3, ..., x_n$. Hovoříme tedy o četnosti hodnot sledovaného znaku. Součet četností všech možných n hodnot znaku se rovná počtu všech jednotek souboru. (Calda a Dupač 1993)

2.2.1 Absolutní četnost

Absolutní četnost nebo jen četnost hodnoty x_j je počet výskytů této hodnoty ve sledovaném souboru. Četnost hodnoty x_j označujeme n_j, kde $j \epsilon \{1, 2, 3, ..., k\}$.

$$n_1 + n_2 + n_3 + ... + n_k = \sum_{j=1}^{k} n_j = n$$

Příklad 1

V 5. třídě je celkem 25 žáků. V následující tabulce máme zaznamenáno, kolik žáků dostalo na vysvědčení z matematiky známku 1, 2, 3, 4, 5.

Tabulka č. 1 – Tabulka dat k příkladu 1

známka	počet žáků (Absolutní četnost)
Výborně	6
Chvalitebně	9
Dobře	5
Dostatečně	3
Nedostatečně	2

Z tabulky výše čteme jednotlivé absolutní četnosti.

Pokud si předdefinujeme sestupně hodnoty známek $n_1 - n_5$, dostaneme postupně hodnoty.

$$n_1 = 6$$
$$n_2 = 9$$
$$n_3 = 5$$
$$n_4 = 3$$
$$n_5 = 2$$
$$n_1 + n_2 + n_3 + n_4 + n_5 = n$$
$$6 + 9 + 5 + 3 + 2 = 25$$

2.2.2 Relativní četnost

Relativní četnost značí, jaká část souboru má hodnotu znaku x_j^*. Kde četnost hodnoty x_j označujeme n_j, kde j$\epsilon\{1, 2, 3, \ldots, k\}$

$$v_j = \frac{n_j}{n}$$

Součet relativních četností je rovna jedné.

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

Relativní četnost lze vyjádřit také v procentech, jejich součet je pak 100%.

Pokud nás nezajímá počet žáků, kteří jakou známku dostali, ale chceme vědět procentuální zastoupení známek žáků, poslouží nám **relativní četnost**.

Tabulka č. 2 – Tabulka četností (příklad 1)

známka	počet žáků (Absolutní četnost)	Relativní četnost	Relativní četnost v %
Výborně	6	0,24	24%
Chvalitebně	9	0,36	36%
Dobře	5	0,2	20%
Dostatečně	3	0,12	12%
Nedostatečně	2	0,08	8%

Máme 25 žáků. Počet označíme; n=25. Relativní četnost dostaneme tak, že absolutní četnost vydělíme n. Z předchozí tabulky vidíme, že známku výborně dostalo 6 žáků. Relativní četnost tedy vypočítáme 6 ÷ 25. Výsledkem relativní četnosti je 0,24. Pokud chceme pro lepší představivost výsledek v procentech, výsledek vynásobíme 100, tedy 24%.

Pro n_1: $6 \div 25 = 0{,}24$

$0{,}24 \cdot 100 = 24\%$

2.2.3 Kumulativní absolutní četnost

Kumulativní četnost je postupně sčítaná četnost jednotlivých vzestupně uspořádaných hodnot statistického znaku ve statistickém souboru.

Na prvním řádku je kumulativní četnost rovna absolutní četnosti a na posledním řádku musí být kumulativní četnost shodná s celkovým počtem jednotek.

Pro druhý znak jsme kumulativní četnost získali tak, že jsme sečetli absolutní četnost prvního a druhého řádku, tedy 6 + 9. Ve třetím řádku je kumulativní četnost 20, tedy 6 + 9 +5, atd.

Tabulka č. 3 – Tabulka kumulativních absolutních četností

známka	počet žáků (Absolutní četnost)	Kumulativní absolutní četnost
Výborně	6	6
Chvalitebně	9	15
Dobře	5	20
Dostatečně	3	23
Nedostatečně	2	25

2.2.4 Kumulativní relativní četnost

Kumulativní četnost lze vyjádřit také relativně. Opět v prvním řádku musíme mít kumulativní relativní četnost shodnou s relativní četností a postupně přičítáme jednotlivé kumulativní relativní četnosti. V poslední řádku musí být kumulativní relativní četnost rovna 1, tedy 100%.

Tabulka č. 4 – Tabulka kumulativních relativních četností

známka	**Relativní četnost**	**Kumulativní relativní četnost**
Výborně	0,24	0,24
Chvalitebně	0,36	0,6
Dobře	0,2	0,8
Dostatečně	0,12	0,92
Nedostatečně	0,08	1

Nyní můžeme do jedné tabulky shrnout všechny četnosti.

Tabulka č. 5 – Souhrnná tabulka četností

	Četnost		**Kumulativní četnost**	
známka	absolutně	Relativně v %	Absolutně	relativně v %
Výborně	6	24%	6	24%
Chvalitebně	9	36%	15	60%
Dobře	5	20%	20	80%
Dostatečně	3	12%	23	92%
Nedostatečně	2	8%	25	100%
Celkem	25	100%		

2.3 Grafické znázornění četností

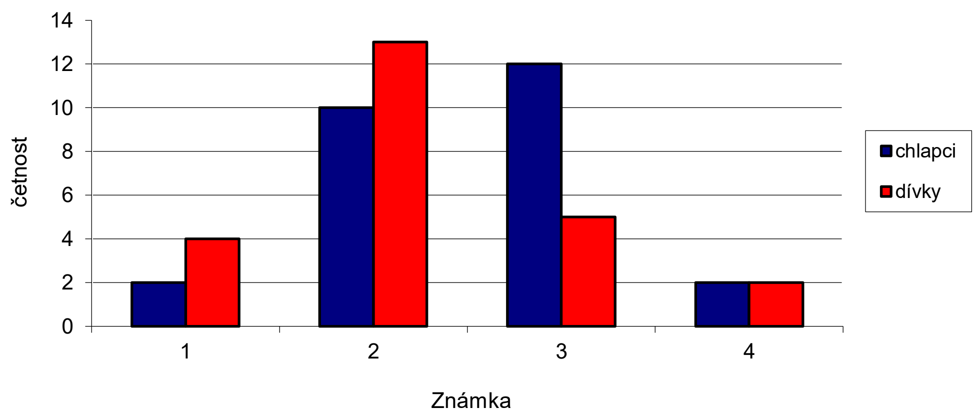
Rozdělení četností lze také znázornit graficky. Mezi tato znázornění řadíme sloupcové diagramy, histogramy, kruhové diagramy, spojnicové diagramy či úsečkové grafy, které si blíže popíšeme v následujících kapitolách.

2.3.1 Sloupcový diagram

Sloupcový graf nebo sloupcový diagram je diagram, který znázorňuje složení sledovaného souboru pomocí obdélníkových pruhů, jejichž výška odpovídá velikosti hodnot, které znázorňují. Graf může být nakreslený svisle i vodorovně. Sloupcový graf poskytuje rychlý přehled o poměrech jednotlivých hodnot. Při sestrojování sloupcového grafu je důležité zvolit si vhodné měřítko. Poté si pojmenujeme osy podle sloupců tabulky a označíme stupnicemi odpovídajícími datům v tabulce. Informace získáme z výšek sloupců. Přesnější hodnotu pak odečteme ze svislé osy grafu.

Prostorový neboli trojrozměrný sloupcový graf, obr. 8, je sice vizuálně působivější, ale může být zavádějící. Díky perspektivě se zdá, že horní plochy sloupců vyjadřují dvě hodnoty četnosti. Avšak skutečnou četnost ukazuje přední plocha sloupce (Vordermanová 2015).

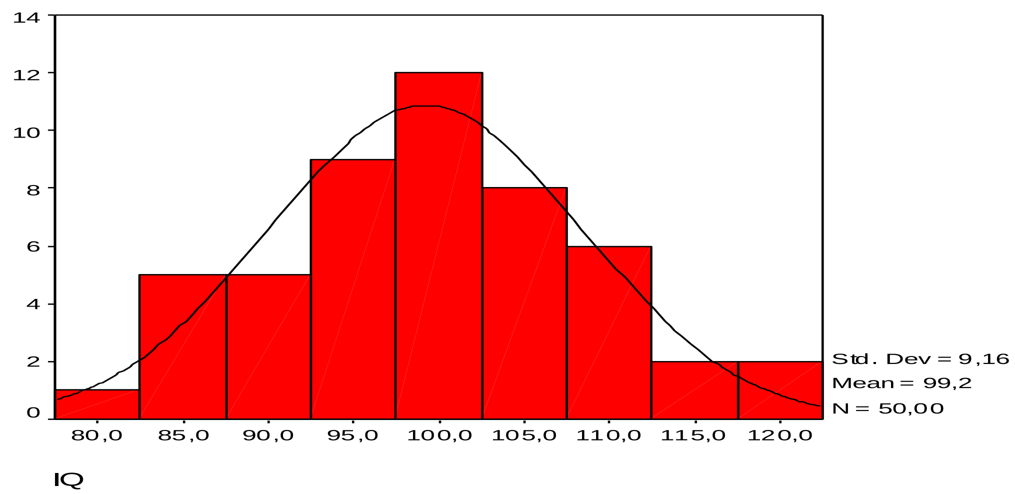
Obrázek č. 1



2.3.2 Histogram

Histogram je grafické znázornění distribuce dat pomocí mnohoúhelníku (obrys sloupců) se sloupci stejné šířky, vyjadřující šířku intervalů (tříd), přičemž výška sloupců vyjadřuje četnost sledované veličiny v daném intervalu, čase. Je důležité zvolit správnou šířku intervalu, neboť nesprávná šířka intervalu může snížit informační hodnotu diagramu.

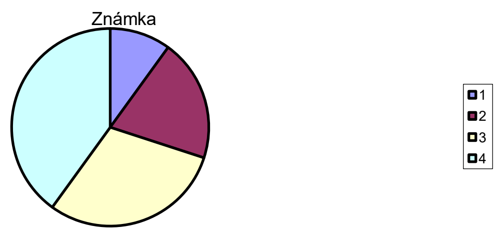
Obrázek č. 2



2.3.3 Kruhový diagram

Kruhový (výsečový nebo koláčový) diagram nebo graf je způsob grafického znázornění struktury sledovaného souboru. Výsečové grafy znázorňují data pomocí kruhových výsečí, kde každá výseč reprezentuje příslušnou třídu souboru dat. Plocha kruhu představuje celý soubor. Výsečové grafy jsou velmi často používány pro svůj okamžitý vizuální účinek. Velikost každé z výsečí jasně prozrazuje poměrné velikosti jednotlivých tříd statistického souboru, což znamená, že údaje lze porovnat snadno a rychle. (Vordermanová 2015).

Obrázek č. 3



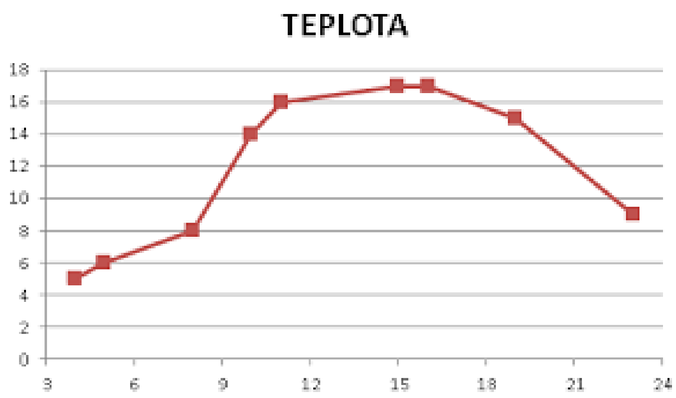
2.3.4 Spojnicový diagram

Spojnicový graf znázorňuje data lomenou čárou tvořenou úsečkami. Spojnicový graf je jedním ze způsobů přesné a přehledné prezentace dat. Využívá se zejména k znázornění dat v časovém období. Třídy jsou na grafu vyznačeny na ose x a četnost na ose y. Jednotlivé body představují četnosti jednotlivých tříd a jednotlivé úsečky představují trendy (Vordermanová 2015)

Spojnicový diagram získáme spojením bodů, jejichž první (x-ová) souřadnice je hodnota znaku, resp. středu intervalu a druhá (y-ová) souřadnice odpovídající četnost.

V praxi toto zobrazení využíváme například ke sledování vývoje hodnoty měny, vývoj teploty vzduchu a další.

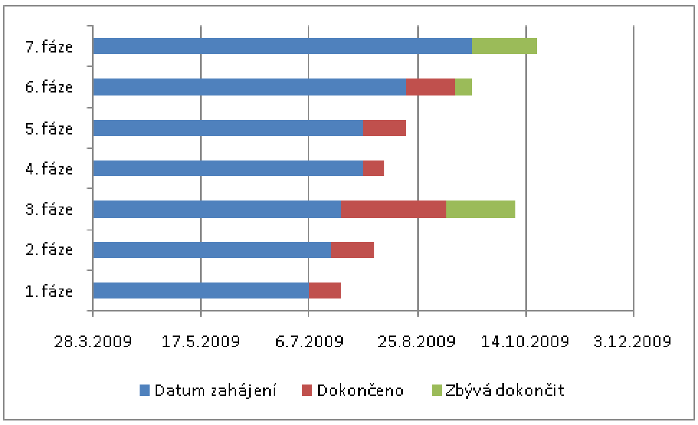
Obrázek č. 4



2.3.5 Úsečkový graf

Úsečkové grafy představují techniku diagramů, které velmi jednoduše a názorně ukazují sled úkolů a jejich začátky a konce. často také bývají využívány pro měření realizace prací a úkolů.

Obrázek č. 4



3. Charakteristiky polohy znaku

Míry polohy vyjadřují, kde (popř. kolem jakého čísla) se data nacházejí; charakterizují „střed“ datového souboru, kolem něhož hodnoty kolísají. Charakteristiky polohy výběrového souboru často odhadují skutečnou střední hodnotu popisované náhodné veličiny. Průměr představuje „střední“ hodnotu souboru metrických dat. Jedná se o typickou hodnotu, která reprezentuje celý soubor. Ve statistice slovo průměr uslyšíte nejčastěji. Máme několik druhů průměrů, například aritmetický průměr, geometrický průměr, vážený průměr a harmonický průměr. Každý z nich se používá na jiný typ dat. V některých případech jsou vhodné doplňkové průměrné hodnoty – modus a medián. V běžném životě se s ním setkáváme dnes a denně, především se jím rozumí aritmetický průměr sledovaných hodnot. Průměrné hodnoty znaku se též nazývají charakteristiky polohy znaku. (Hazardní hry 2007).

3.1 Aritmetický průměr

Aritmetický průměr je nejčastěji užívanou charakteristiku polohy znaku x. S aritmetickým průměrem se setkáváme v běžném životě nejčastěji. Dalo by se říci, že téměř denně. Zajímáme se o průměrný plat, průměrnou výšku, váhu, průměrný prodej nejrůznějších výrobků a děti ve škole si vypočítávají průměrnou známkū z daného předmětu. Aritmetický průměr se značí $\bar{x}$ a je dán vzorcem, kde x_i značí součet všech hodnot x_1, x_2,.., x_n.

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

Součet zjištěných hodnot znaku u všech jednotek souboru vydělený počtem všech jednotek souboru (Polák 2014; Calda a Dupač 1993).

Vlastnosti aritmetického průměru:

• přičtením, odečtením, vynásobením nebo vydělením všech hodnot znaku nenulovým číslem se odpovídajícím způsobem změní také aritmetický průměr (např. zvětšíme-li všechny hodnoty o 2, zvětší se aritmetický průměr také o 2);

• rozdělíme-li soubor do skupin, pak průměr celého souboru je váženým průměrem skupinových průměrů, přičemž jako váhy vystupují počty jednotek v jednotlivých skupinách.

Příklad 1

V tabulce níže, máme naměřeny výšky náhodně změřených lidí. Spočítáme průměrnou výšku statistického souboru.

Tabulka č. 6

Výška	160	165	170	175
Počet	1	2	1	2

Řešení

1. Seřadíme všechny hodnoty podle velikosti.

160, 165, 165, 170, 175, 175

2. Sečteme všechny hodnoty v tomto souboru.

160 + 165 + 165 + 170 + 175 +175 = **1010**

3. Vydělíme tento součet počtem hodnot.

$1010 \div 6 = 168{,}33$

$\bar{x}$ = 168,33

Aritmetický průměr výšek osob je 168,33 cm.

Příklad 2

Nyní spočítáme aritmetický průměr výšek, naměřených u žáku 8. třídy.

Tabulka č. 7

Osoba	Anna	Petra	Adam	Jan	Tomáš
výška	145	160	165	170	170

$$\bar{x} = \frac{145 + 160 + 165 + 170 + 170}{5}$$

$$\bar{x} = 162$$

Příklady k procvičení

1) Při měření rozlohy bytů jsme naměřili následující hodnoty v m^2:
82,6; 57,3; 70,4; 65; 48,4; 103,8; 73,6; 43,5; 66,1; 93; 52,6; 70; 84,2; 55; 81,3; 61,5; 75,1; 34,8; 62,4; 116; 70,1; 63,6; 93; 59,2; 65,9; 77,2; 52,8; 68,7; 79,2; 87,4.
Spočítej průměrnou rozlohu bytu.

2) Pelikánovi dávali na účet průběžně tyto částky:
2 000, 4 500, 6 100, 2 100, 3 000, 1 700, 2 600, 2 000
Urči průměrné úložné částky.

3) Při zjišťování velikosti nohy mezi deseti dívkami v kvartě
jsme získali tyto údaje (v cm): 27, 24, 25, 25, 24, 23, 25, 26, 27, 23.
Určete průměrnou délku chodidla.

4) K fotbalovém zápasu nastoupí 11 hráčů. Jejich průměrný věk je 25 let. Během zápasu je kvůli zranění jeden obránce vystřídán sedmnáctiletým hráčem a průměrný věk mužstva je pak 23. Jak starý byl zraněný hráč?

3.2 Geometrický průměr

Geometrický průměr $\overline{x_G}$ zavádíme jen pro kladná čísla na rozdíl od aritmetického průměru. V praxi se například používá k určení průměrného ročního tempa růstu. Pokud bychom si jednotlivá období očíslovali 0,1,2, ..., n, pak jim odpovídající hodnoty znaku jsou x_0, x_1, x_2, ..., x_n a přírůstky za jednotlivá období jsou

$y_1 = x_1 - x_0, y_2 = x_2 - x_1, ..., y_n = x_n - x_{n-1}$.

Průměrný přírůstek je pak roven $\overline{y} = \frac{x_{n} - x_0}{n}$.

Pokud se ptáme na průměrné tempo růstu za jedno období, myslí se tím průměr podílů hodnot za dvě po sobě následující období, tedy $g_1 = \frac{x_1}{x_0}, ..., g_n = \frac{x_n}{x_{n-1}}$.

Geometrický průměr pak spočítáme

$$\overline{x_G} = \sqrt[n]{g_1 \cdot g_2 \cdot ... \cdot g_n} = \sqrt[n]{\frac{x_n}{x_0}}$$

Všechny hodnoty mezi sebou vynásobíme a uděláme n-tou odmocninu (Polák 2014; Calda a Dupač 1993).

3.2.1 Vážený geometrický průměr

Vážený geometrický průměr, kdy prvek x_1 má četnost výskytu n_1, prvek x_2 má četnost výskytu n_2, atd. až prvek x_k má četnost výskytu n_k, je pak

$$x_g = \sqrt[n]{{x_1}^{n_1} \cdot {x_2}^{n_2} \cdot ... \cdot {x_k}^{n_k}}$$

Kde $n = n_1 + n_2 + \cdots + n_k$

Příklad 1

Obchodník prodával jeden kus zboží za 100 Kč. Rozhodl se jej zdražit o 20 % na 120 % hodnoty. Následně obchodník zdražil zboží o dalších 30 % na 130 % z již zvýšené hodnoty. O kolik průměrně obchodník zdražil zboží při jednom zdražení?

Řešení

Na začátek spočítáme kolik zboží stálo po celkovém zdražení. Tato hodnota nám následně bude sloužit pro kontrolu, zda jsme počítali správně. Cena zboží po prvním zdražení byla $100 \cdot 1{,}2 = 120$ Kč. Po druhém zdražení zboží stálo $120 \cdot 1{,}3 = 156$ Kč.

Vidíme, že první koeficient růstu byl 1,2 a druhý koeficient růstu byl 1,3. My vypočítáme průměrný koeficient.

Sestavíme rovnici $100 \cdot x \cdot x = 156$. Neznámá x je námi hledaný průměrný koeficient. Rovnici nyní vyřešíme.

$$100 \cdot x \cdot x = 156$$

$$x^2 = \frac{156}{100}$$

$$x = \pm\sqrt{\frac{156}{100}}$$

$$x = \pm\sqrt{1,56}$$

Jelikož počítáme růst, pak budeme uvažovat jen kladnou hodnotu a to $\sqrt{1,56}$.

Průměrný koeficient růstu je tedy $x = 1,249$

Výpočtem pro ověření $100 \cdot \sqrt{1,56} \cdot \sqrt{1,56} = 156$ zjistíme, že jsme počítali správně.

Příklad 2

Přírůstek maximálního denního kurzu akcií Českého Telecomu v RM-Systému ze 16. na 17. ledna 2006 byl 1,10777 %, ze 17. na 18. ledna 2006 -0,79852 % a z 18. na 19. ledna 2006 1,08573 %. Jaký byl průměrný denní přírůstek maximální ceny 1 akcie v RM-Systému v těchto třech obdobích?

Denní nárůstky:

1. den 101,10777%

2. den 99,20148%

3. den 101,08573%

$$x = \sqrt[3]{101,10777 \cdot 99,20148 \cdot 101,08573}$$

$$x = 100,4610034$$

Výsledek: 0,4610034 %

Příklady k procvičení

1) Při laboratorním cvičení z fyziky jsme měřili délku válečku. Zjištěné hodnoty jsou uvedené v milimetrech: 102; 99; 103; 105; 98; 99; 100; 102; 100; 106.

Vypočítejte geometrický průměr.

2) Výroba se v prvním roce zvýšila 2krát, ve druhém roce se proti předchozímu roku zvýšila 6krát.

a) Kolikrát se zvýšila celkem ?

b) Kolikrát se zvýšila průměrně ročně?

3) Zkoumáme inflaci v ČR v letech 2001-2011, z českého statistického úřadu jsme zjistili tyto údaje: 4,7 1,8 0,1 2,8 1,9 2,5 2,8 6,3 1,0 1,5 1,9 Vypočítejte geometrický průměr.

4) Vypočtěte průměrný růst cen za poslední čtyři roky, jestliže byl zaznamenán postupný nárůst o 20 % z předchozího roku a pak 10 %, poté 15 % pokles cen, ale pak zase 10 % růst cen.

3.3 Harmonický průměr

Harmonický průměr kladných hodnot statistického souboru je definován jako převrácená hodnota aritmetického průměru převrácených hodnot, tzn. podíl počtu členů a součtu převrácených hodnot zadaných členů. Harmonický průměr se využívá k zjištění průměrné délky času potřebné k provedení nějakého úkonu, který provádí současně několik osob či strojů. Například při výpočtu průměrné rychlosti na úsecích stejné délky.

Výsledek harmonického průměru je vždy menší nebo roven geometrickému průměru (Polák 2014; Calda a Dupač 1993).

Vzorec harmonického průměru

$$\overline{x_H} = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + .. + \frac{1}{x_n}\right)}$$

Příklad 1

Auto jelo první polovinu cesty průměrnou rychlostí v_1 = 20 km/h a druhou polovinu cesty průměrnou rychlostí v_2 = 80 km/h. Jakou průměrnou rychlostí auto jelo?

Řešení:

Auto každou polovinu cesty jelo různou rychlostí, proto je ujelo za různé časy.

Průměrnou rychlost auta vypočítáme jako harmonický průměr rychlostí v_1 a v_2.

$$\bar{v} = \frac{n}{\frac{1}{v_1} + \frac{1}{v_2}}$$

$$\bar{v} = \frac{2}{\frac{1}{20} + \frac{1}{80}} = \frac{2}{\frac{4+1}{80}} = \frac{2 \cdot 80}{5} = \frac{160}{5} = 32\ km/h$$

Průměrná rychlost auta byla 32 km/hod.

Příklad 2

V kavárně jsou provozovány dva kávovary současně. Příprava kávy na starším stroji zabere 3 minuty, zatímco na novějším stroji pouze 1,5 minuty. Jak dlouho trvá v průměru příprava jednoho šálku kávy?

$$\bar{v} = \frac{n}{\frac{1}{v_1} + \frac{1}{v_2}}$$

$$\bar{v} = \frac{2}{\frac{1}{3} + \frac{1}{1,5}} = \frac{2}{\frac{1+2}{3}} = \frac{2}{1} = 2\ minuty$$

Výsledek: 2 minuty

Příklady k procvičení

1) Žaneta nasbírá kilogram ovoce za 2,25 hodiny. Ondřej nasbírá kilogram ovoce za 1,5 hodiny. Za jak dlouho nasbírají společně kilogram ovoce?

2) Řidič zkušebního automobilu jel do cílového místa rychlostí 70 km/h a zpět rychlostí 90 km/h. Jakou průměrnou rychlost dosáhl na celé trase?

3) Tři pracovníci opakovaně provádějí stejnou operaci. Prvnímu trvá operace 2 minuty, druhému 3 minuty, třetímu 4 minuty. Jak dlouho trvá průměrně jedna operace?

4) Vypočtěte celkové průměrné rychlosti dojíždějících do centra. Když víte, že dostupnost z místa A = 30 min., z bodu B = 20 min. a z bodu C = 6 min.

3.4 Modus

Modus je další charakteristika polohy, která se používá pro číselná i nečíselná data, u kterých nelze spočítat průměr. Modus je hodnota znaku, která má v souboru největší zastoupení, neboli má v souboru největší četnost. Značíme jej *Mod(x)* (Polák 2014; Calda a Dupač 1993).

Se stejnou četností se může vyskytnout více hodnot, tzn., že modus nemusí být vždy jednoznačně určen.

Příklad 1

Student získal během měsíce října tyto známky: 5; 3; 2; 1; 3; 2; 4; 2; 2; 3; 3; 1. Určete modus.

Řešení: Největší zastoupení ve výčtu prvků je hodnota 2 a 3.

Příklad 2

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.

Řešení: Modus je 175 cm.

3.5 Medián

Medián dělí soubor podle velikostí seřazených číselných hodnot znaku na dvě poloviny. Je to tedy číslo, které se nachází přesně uprostřed číselné řady. Medián je prostřední hodnota znaku, pokud máme soubor, v němž jsou hodnoty znaku uspořádány vzestupně, platí

$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

Medián značíme Med(x) nebo $\tilde{x}$

Mylně si mnoho lidí myslí, že aritmetický průměr je číslo, které je přesně uprostřed souboru dat, prostřední číslo je ovšem medián. Výhoda mediánu je, že není ovlivněn extrémními hodnotami. Proto se často používá tam, kde aritmetický průměr nedává vhodné výsledky (Polák 2014; Horenský a kol. 2015; Calda a Dupač 1993).

Seřadíme a vybereme prostřední člen. Musí platit, že 50% dat je větších nebo rovno než medián a 50% dat je menších nebo rovno.

Při sudém počtu vzorků spočítáme průměr dvou prostředních čísel.

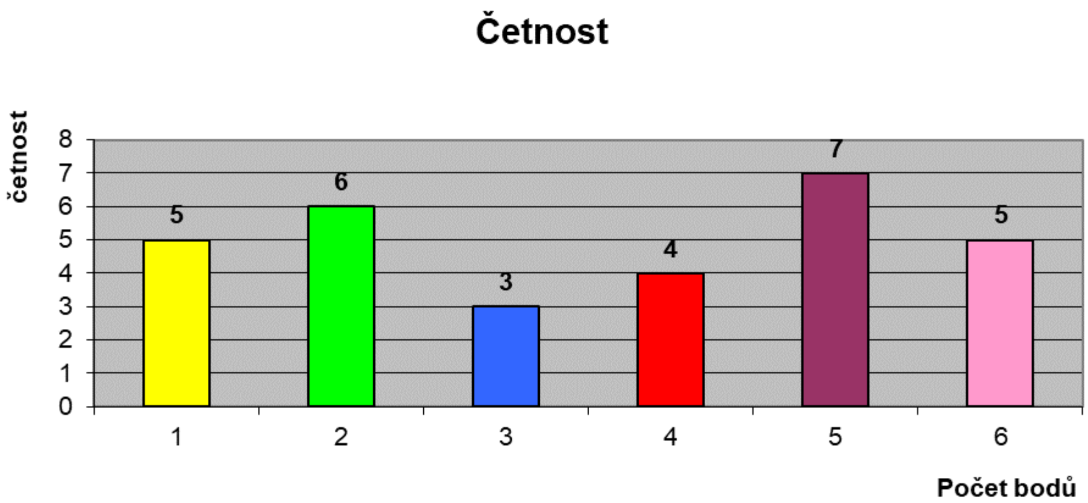
Jako dobrý příklad máme hodnocení prospěchu. Známka dobře je sice prostřední sloupec tabulky, ale neurčuje střední hodnotu souboru. Musíme zohlednit četnost jednotlivých známek (Polák 2014; Horenský a kol. 2015; Calda a Dupač 1993).

Příklad 1

Tonda házel 30x s hrací kostkou a postupně dosáhl těchto hodnot:

4, 6, 2, 5, 5, 2, 2, 5, 1, 6, 5, 3, 5, 2, 6, 1, 6, 4, 6, 5, 2, 3, 1, 5, 3, 4, 4, 1, 1, 2

Zjisti medián výčtu prvků:



Seřadíme hodnoty od nejmenšího po největší a vybereme prostřední hodnotu.

1,1,1,1,1,2,2,2,2,2,2,3,3,3,4,4,4,4,5,5,5,5,5,5,5,6,6,6,6,6

Řešení: Prostřední hodnota, také jinak 15. a 16. znak má hodnotu 4 body.

$$\tilde{x} = 4$$

Příklad 2

Následující data představují velikosti triček prodaných při výprodeji firmy TRIKO.

S, M, L, S, M, L, XL, XL, M, XL, XL, L, M, S, M, L, L, XL, XL, XL, L, M

Zjisti medián výčtu prvků:

Nejprve seřadíme velikosti od S do XL, abychom dostali ordinální data.

S,S,S,M,M,M,M,M,M,L,L,L,L,L,L,XL,XL,XL,XL,XL,XL,XL

S – 3x

M – 6x

L – 6x

XL – 7x

Celkem máme 22 kusů oblečení, mediánem bude tedy 11. a 12. prvek.

S,S,S,M,M,M,M,M,M,L,L,L,L,L,L,XL,XL,XL,XL,XL,XL,XL

Řešení: L

3.6 Kvartily

Kromě mediánu máme ještě kvartily. Dolní kvartil, Q1, je hodnota, která z hodnot odděluje nejnižší čtvrtinu hodnot. Například student, který je na úrovni dolního kvartilu, tedy ví, že přesně čtvrtina studentů je na tom hůře, a přesně tři čtvrtiny lépe. V případě, že výsledky studentů seřadíme od nejhoršího po nejlepší. Obdobně je to samozřejmě s horním kvartilem. Pak máme ještě kvartil Q2, který odděluje obě poloviny souboru - ale tomu se, jak už víme, říká medián. Druhý kvartil a medián jsou tedy stejné hodnoty.

Výpočet kvartilu je následující. Setřídíme soubor dat s n prvky od nejmenšího po největší a vybereme n/4-tý prvek jako kvartil Q1 a 3n/4-tý prvek jako kvartil Q3. V případě lichého počtu n prvků je Q1 či Q3 průměrem 2 sousedících prvků na rozmezí kvartilů.

Například pro 1000 prvků, je první kvartil Q1 rovný 250-tému prvku setříděných dat a kvartil Q3 rovný 750-tému prvku setříděných dat.

Příklad 1

Následující data představují věk hudebníků vystupujících na přehlídce dechových orchestrů. Vypočti dolní a horní kvantil:

22, 82, 27, 43, 19, 47, 41, 34, 34, 42, 35.

Řešení:

Seřazená data: 19, 22, 27, 34, 34, 35, 41, 42, 43, 47, 82.

Q1: p = 0, 25; n = 11 ⇒ zp = 11 · 0, 25 + 0, 5 = 3, 25.

Q1 je tedy průměrem prvků s pořadím 3 a 4.

Q1 = $\frac{27+34}{2}$

Q1 = 30, 5 let,

25% hudebníků vystupujících na přehlídce dechových orchestrů je mladších než 30,5 let (75% z nich má 30,5 let a více).

Q3: p = 0, 75; n = 11 ⇒ zp = 11 · 0, 75 + 0, 5 = 8, 75

Q3 je tedy průměrem prvků s pořadím 8 a 9.

Q3 = $\frac{42+43}{2}$

= 42, 5 let

75% hudebníků vystupujících na přehlídce dechových orchestrů je mladších než 42,5 let (25% z nich má 42,5 let a více).

Příklad 2

Z datového souboru 1,4 9,9 0,2 9,9 9,6 4,1 2,3 0,9 4,8 7,6 1,9 1,0 3,1 8,1 4,5 3,9 0,3 2,8 0,5 3,6 vypočtěte Q1 a Q3.

Opět seřadíme data od nejmenšího po největší.

0,2 0,3 0,5 0,9 1,0 1,4 1,9 2,3 2,8 3,1 3,6 3,9 4,1 4,5 4,8 7,6 8,1 9,6 9,9 9,9

Máme 22 prvků

Q1: p = 0, 25; n = 20 ⇒ zp = 20 · 0, 25 = 5.

Q1 je tedy průměrem prvků s pořadím 5 a 6.

Q1 = $\frac{1,0+1,4}{2}$

Q1 = 1,2

Q3: p = 0, 75; n = 20 ⇒ zp = 20 · 0, 75 = 15

Q3 je tedy průměrem prvků s pořadím 15 a 16.

Q3 = $\frac{4,8+7,6}{2}$

Q3 = 6,2

Řešení:

Q1 = 1,2

Q3 = 6,2

4. Charakteristiky variability

Charakteristiku polohy označujeme číslem, kolem kterého jednotlivé hodnoty znaku kolísají, např. aritmetický průměr, modus a medián. Velikost tohoto kolísání vyjadřují charakteristiky proměnlivosti znaku. Například kdybychom zkusili zvážit několik sáčků mouky, která má vážit 1kg, zjistili bychom, že každý sáček váží jinak, mají odchylku od určité hodnoty (Horenský a kol. 2015; Hazardní hry 2007).

4.1. Odchylka

4.1.1. Relativní odchylka

Relativní odchylka také používaná v praxi nám určuje podíl průměrné absolutní odchylky a příslušného aritmetického průměru. Vyjadřuje se většinou v procentech. Značíme ji r. Jedná se o číslo, které nemá jednotku a tak nám umožňuje porovnání souborů, v nichž mají hodnoty znaků různé jednotky, např. měření v milimetrech a metrech. Tedy

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

Ve fyzice se pro relativní odchylku používá termín relativní chyba. Relativní chyba měření vyjadřuje přesnost měření. Laboratorní měření ve školních podmínkách považujeme za přesné, pokud je relativní chyba menší než 1% (Horenský a kol. 2015).

Příklad

Byly zjištěny následující hodnoty: 10 , 12 , 15 , 10 , 8 , 17, 12 , 11

Průměrná hodnota: $\bar{x} = \frac{95}{8}$

$\bar{x} = 11{,}875$

Absolutní odchylka pro první člen je 11,875 – 10 = 1,875

Relativní odchylku spočítáme takto: $(1{,}875 \div 11{,}875) \cdot 100 = 15{,}79\%$

4.1.2. Kvartilová odchylka

Kvartilová odchylka udává průměrnou vzdálenost mezi dvěma kvartily. Jinými slovy je definovaná jako polovina kvartilového rozpětí.

Značí se $QD = \frac{Q3-Q1}{2}$ (Horenský a kol. 2015)

Příklad

Zadání z kapitoly 3.6 příklad 1.

$$Q1 = 30{,}5$$

$$Q3 = 42{,}5$$

$$QD = \frac{30{,}5 + 42{,}5}{2}$$

$$QD = 36{,}5$$

4.1.3. Směrodatná odchylka

Směrodatná odchylka představuje nejčastější charakteristiku variability. V praxi se s ní setkáváme častěji než s rozptylem a je definována jako kvadratický průměr odchylek hodnot znaku od jejich aritmetického průměru. Směrodatná odchylka je vyjádřena ve stejných jednotkách jako sledovaný znak, např. pokud jsou metrická data v jednotkách *m*, pak výsledek směrodatné odchylky bude v jednotkách m^2. Jednoduše řečeno je směrodatná odchylka druhá odmocnina z rozptylu. Značíme ji s_x (Horenský a kol. 2015; Calda a Dupač 1993).

Směrodatná odchylka vyjadřuje, jak jsou jednotlivé hodnoty vyrovnané, jak jsou vzdálené od průměru. Druhou odmocninu používáme, protože rozptyl nemá stejnou jednotku jako měřená veličina (Calda a Dupač 1993; Horenský a kol. 2015).

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x})^2}$$

Příklad

Najděte směrodatnou odchylku pro množinu dat:

Věk	Počet osob
0-10	15
10-20	15
20-30	23
30-40	22
40-50	25
50-60	10
60-70	5
70-80	10

$$n = 15 + 15 + 23 + 22 + 25 + 10 + 5 + 10 = 125$$

$$m_1 = 5 \cdot 15 + 15 \cdot 15 + 25 \cdot 23 + 35 \cdot 22 + 45 \cdot 25 + 55 \cdot 10 + 65 \cdot 5 + 75 \cdot 10 = 4395$$

$$m = \frac{m_1}{n} = \frac{4395}{125} = 35{,}16$$

$$s_1 = 5^2 \cdot 15 + 15^2 \cdot 15 + 25^2 \cdot 23 + 35^2 \cdot 22 + 45^2 \cdot 25 + 55^2 \cdot 10 + 65^2 \cdot 5 + 75^2 \cdot 10 = 203325$$

$$s_2 = \frac{s_1}{n} - m^2 = \frac{203325}{125} - 35{,}16^2 = 390{,}3744$$

$$s = \sqrt{s_2} = \sqrt{390{,}3744} = 19{,}7579$$

4.1.4. Variační koeficient

Posledním ze základních ukazatelů variability je variační koeficient, který vypovídá o relativním významu průměrné odchylky od průměru, tj. kolik procent průměru představuje směrodatná odchylka. Variační koeficient vypočítáme podílem směrodatné odchylky a aritmetického průměru. Tento ukazatel je nejvhodnější pro porovnání variability ukazatelů a souborů jednotek různých úrovní, neboť jde o bezrozměrnou veličinu obvykle vyjádřenou v procentech.

$$v_x = \frac{s_x}{\bar{x}}$$

Příklad

Tabulka uvádí naměřené rozdíly v počtu kusů v obalu oproti počtu kusů uvedenému na obalu. Určete variační koeficient.

Rozdíl	Počet měření
-2	10
-1	4
0	5
1	3
2	8

Nejprve určíme průměr.

$$\bar{x} = \frac{-2 \cdot 10 + 1 \cdot 4 + 0 \cdot 5 + 1 \cdot 3 + 2 \cdot 8}{10 + 4 + 5 + 3 + 8} = 0{,}1$$

Vypočtený průměr využijeme k výpočtu rozptylu

$$s_x^2 = \frac{(-2 - 0{,}1)^2 \cdot 10 + (-1 - 0{,}1)^2 \cdot 4 + (0 - 0{,}1)^2 \cdot 5 + (1 - 0{,}1)^2 \cdot 3 + (2 - 0{,}1)^2 \cdot 8}{10 + 4 + 5 + 3 + 8} = 2{,}6233$$

Odmocněním rozptylu získáme směrodatnou odchylku

$$s_x = 1{,}6197$$

Nyní lze určit variační koeficient

$$v_x = \frac{1{,}6197}{0{,}1} = 16{,}197$$

Závěr

Cílem mé bakalářské práce bylo sestavit ucelený a přehledný učební materiál pro žáky 2. stupně základní školy a studenty střední školy ze základů statistiky.

V úvodu práce jsme si definovali pojmy týkající se statistiky, co jsou data, soubor, jednotka, či znak. Poté následovaly témata týkající se zpracování statistických dat, jejich vysvětlení a řešený vzorový příklad.

Seznámili jsme se s tématy, jako jsou různé druhy četností, aritmetický, geometrický či harmonický průměr. Modus, medián, kvartily a dále odchylky nebo variační koeficient.

Danou prací bych se rád dále zabýval i ve své diplomové práci, kde by práce byla rozvinuta o didaktický materiál pracovních listů s příklady ze statistiky.

Seznam použité literatury

CALDA, Emil a Václav DUPAČ. Matematika pro gymnázia. 5. vyd. Praha: Prometheus, 2008, 170 s. Učebnice pro střední školy (Prometheus). ISBN 978-80-7196-365-3.

ČERMÁK, Pavel a Petra ČERVINKOVÁ. Odmaturuj! z matematiky: přehled středoškolského učiva. Vyd. 3. opr. Brno: Didaktis, 2004, 208 s. Odmaturuj!. ISBN 80-735-8014-4

KUBEŠOVÁ, Naděžda a Eva CIBULKOVÁ. Matematika: přehled středoškolského učiva. 1. vyd. Třebíč: Petra Velanová, 2006, 239 s. Maturita (Petra Velanová). ISBN 80-868-7303-X

HORENSKÝ, Radek, a kol. Matematika pro střední školy 8. díl: Kombinatorika, pravděpodobnost, statistika. Brno: Didaktis, 2015.

POLÁK, Josef. Didaktika matematiky: Jak učit matematiku zajímavě a užitečně. Plzeň: Fraus, 2014.

ŘEZANKOVÁ, H., L. MAREK a M. VRABEC. Interaktivní učebnice statistiky: Typy proměnných [online]. 2001 [cit.2017-05-22]. Dostupné z: http://iastat.vse.cz/typy_promennych.html

VORDERMANOVÁ, Carol. Matematika: Spolu to zvládneme. Praha: Slovart, 2015.

PAVELKA, Jindřich. Hazardní hry [online]. 2007-2017 [cit. 2017-6-20]. Dostupné z: http://www.hazardni-hry.eu/statistika/

ZWERENZ, Karlheinz. Statistik: Einführung in die computergestützte dateanalyse. München: Oldenbourg, 2015