

MORAVSKÁ VYSOKÁ ŠKOLA OLMOUC

Ústav informatiky a aplikované matematiky

Markéta Kristková

Využití software Mathematica ve statistice

Software Mathematica and Statistics

Bakalářská práce

Vedoucí práce: Mgr. Veronika Říhová, Ph.D.

Olomouc 2016

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a použila jen uvedené informační zdroje.

Olomouc.....

Markéta Kristková

Děkuji Mgr. Veronice Říhové, Ph.D. za odborné vedení a užitečné rady, které mi poskytovala po celou dobu tvorby bakalářské práce.

OBSAH

ÚVOD.....	6
1 Představení Softwaru Mathematica	7
1.1 Co je software Mathematica?.....	7
1.2 Vznik	7
1.3 Využití software Mathematica	8
1.4 Uživatelé	8
1.5 Prostředí	9
1.6 Programovací jazyk.....	11
2 Úvod do statistiky	13
2.1 Pojem statistika	13
2.2 Historie a současnost	13
2.3 Základní pojmy	14
3 Charakteristiky statistického souboru	16
3.1 Charakteristiky polohy	16
3.1.1 Aritmetický průměr	17
3.1.2 Geometrický průměr	22
3.1.3 Harmonický průměr.....	26
3.1.4 Modus.....	29
3.1.5 Medián	30
3.2 Charakteristiky variability	33
3.2.1 Variační rozpětí.....	33
3.2.2 Průměrná odchylka.....	34
3.2.3 Rozptyl a směrodatná odchylka	34
3.2.4 Variační koeficient	37
3.3 Charakteristiky šikmosti	39
3.3.1 Koeficient šikmosti	39
3.4 Charakteristiky špičatosti	40
3.4.1 Koeficient špičatosti	40
4 Rozdělení četností a jejich grafické vyjádření	43
4.1 Bodový graf rozložení četností.....	44
4.2 Polygon rozložení četností.....	45
4.3 Úsečkový graf	46
4.4 Sloupcový graf rozložení četností	47
4.5 Histogram rozdělení četností.....	48
4.6 Výsečkový graf	49
5 Praktické využití software Mathematica.....	50

Závěr.....	53
Anotace	54
Literatura a prameny	55
Seznam zkratk	57
Seznam obrázků	58
Seznam tabulek.....	59

ÚVOD

Pro svoji bakalářskou práci jsem si vybrala téma Využití software Mathematica ve statistice. Důvodem, proč jsem si zvolila toto téma, je můj zájem o matematiku a statistiku. Domnívám se, že práce na tomto tématu mi bude přínosná při dalším studiu nebo v praxi, může být nápomocná ostatním studentům vysokých škol nebo každému, kdo se chce naučit pracovat s programem Mathematica.

Cílem této práce je představit program Mathematica a ukázat jak se dá tento program využít pro výpočty z prostředí statistiky, jaké typy příkladů ze statistiky lze pomocí programu řešit a jak jinak můžeme program ve statistice využít. Statistika je obsáhlá vědní disciplína, proto tato práce bude obsahovat jen některá vybraná odvětví. Stejný výběr bude proveden i v programu Mathematica, představíme jen ty funkce a příkazy, které můžeme ve statistice potřebovat.

Myslím si, že toto téma je aktuální. V dnešní době hrají počítačové programy velkou roli, usnadňují lidem práci, jsou přesnější a rychlejší. Díky technice a počítačovým programům můžeme ušetřit náš čas, zdokonaluje své dovednosti, získáváme nové zkušenosti.

Tato práce bude dělena na jednotlivé kapitoly a podkapitoly, nebude zde použito standardní dělení na teoretickou a praktickou část. U každé kapitoly se nejprve zaměříme na teoretické informace a následně budeme řešit příklad v programu Mathematica. Řešení příkladů v programu Mathematica bude vloženo do této práce ve formě obrázků, vždy před každým vloženým obrázkem s řešením bude napsáno detailní vysvětlení řešení příkladu.

Tato práce obsahuje celkem 5 hlavních kapitol, tyto kapitoly se dále člení na podkapitoly. V 1. kapitole se dozvíme informace o samotném softwaru Mathematica. Představíme si jeho prostředí, uživatele, jak vznikl, atd. Ve 2. kapitole se budeme zabývat statistikou, vysvětlíme si základní pojmy a samotný pojem statistika. Ve 3. kapitole se zaměříme na charakteristiky statistického souboru, budou to charakteristiky polohy, variability, šikmosti a špičatosti. Ve 4. kapitole se podíváme na rozdělení četností a jejich grafické vyjádření. A v poslední 5. kapitole se zaměříme na praktické využití software Mathematica.

1 PŘEDSTAVENÍ SOFTWARE MATHEMATICA

V první kapitole se budeme zabývat softwarem Mathematica, jak tento program vznikl, k čemu se program používá, jaké funkce tento program má. Dále popíšeme jeho prostředí a programovací jazyk. V této práci se budeme zabývat software Mathematica verze 10.2.

1.1 Co je software Mathematica?

„Mathematica je počítačový program používaný zejména pro řešení technických a matematických úloh, soubory tohoto programu mají příponu *.nb. Vzhledem k tomu, že matematické výpočty jsou nezbytné prakticky ve všech vědních disciplínách, je program široce využíván prakticky ve všech přírodovědných oborech a v technických vědách, ze společenských věd pak především v ekonomii. Program vyvíjí a prodává firma Wolfram Research se sídlem v Champaign, Illinois v USA, nové verze jsou dostupné pro OS Microsoft Windows, MacOS X a Linux ve třech jazykových mutacích (anglické, čínské a japonské). Kromě standardních verzí je nabízeno velké množství balíčků zaměřených na speciální oblasti, jako například aplikovaná fyzika, finanční matematika, aplikovaná statistika, biomatematika atd.“¹

1.2 Vznik

Stephen Wolfram založil v roce 1987 společnost Wolfram Research, Inc., ve které tento program vznikl. V roce 1988 proběhlo první vydání software Mathematica a toto vydání přispělo k lepšímu využívání počítačů v řadě odvětví a tímto vydání začal nový věk tzv. technical computing. V 60. let minulého století vznikaly tematické sady (numerické, grafické, algebraické), ale byli vhodné vždy pro určitý typ úloh. Až Stephen Wolfram spojil tyto úlohy a integroval je to jednoho výrobku, do software Mathematica. Tak se stal software Mathematica nástroj, který zvládá numerickou a symbolickou matematiku, grafický a dokumentační systém, a je schopný se propojit s dalšími aplikacemi.²

¹ FŇUKAL, M., SMRČKA, D., a KLADIVO P., *Software Mathematica pro geografy*, s. 5.

² Srov. CHRAMCOV, B., *Základy práce v prostředí Mathematica*, s. 7.

1.3 Využití software Mathematica

„Co tedy programový systém Mathematica nabízí svým uživatelům? Mezi základní funkce tohoto programu patří:

- symbolické a numerické výpočty;
- zjednodušování a úpravy algebraických výrazů;
- kreslení grafů a obrazců v rovině i v prostoru;
- výpočty derivací a integrálů;
- řešení rovnic a jejich soustav od jednoduchých až po diferenciální;
- práce s velkými soubory dat a databázemi;
- algoritmičké, funkcionální i logické programování;
- tvorba animací a prezentací (slide show);
- psaní článků a jiných textů s matematickými výpočty.“³

1.4 Uživatelé

Prvotní podnázev programu byl „system for doing mathematics by computer“, což značí, že tento program nemá sloužit pouze odborníkům, ale i široké veřejnosti. Program je uživatelsky velmi přístupný, výrobci se snaží uživatelům pomoci s využíváním všech funkcí programu. Uživatele tohoto programu můžeme najít např. i mezi geografy, které řadíme do kategorie „matematických laiků“. To neznamená, že by tento program byl nějak nezajímavý nebo rutinní pro samotné matematiky, spíše naopak. Kvalita zpracování matematických postupů a algoritmů je na tak vysoké úrovni, že tento program je užitečný i pro ně.⁴

Na počátku se software Mathematica užíval v oblasti fyziky, engineeringu a matematiky, kdy měl přibližně desítky tisíc uživatelů po celém světě. V průběhu několika následujících let se tento program začal uplatňovat v dalších oborech, jako jsou přírodní a biologické vědy. V současné době má program přes jeden milion uživatelů, pracují s ním jak odborníci, tak i studenti, umělci, právníci, skladatelé, vědci.

V praxi tento program můžeme vidět při vývoji a výrobě, při plánování a různých typech analýz, při důležitých objevech, jako nástroj vzdělávání studentů. Je používán jako sofistikovaný nástroj finančního modelování a jako výzkumné a prototypové prostředí.⁵

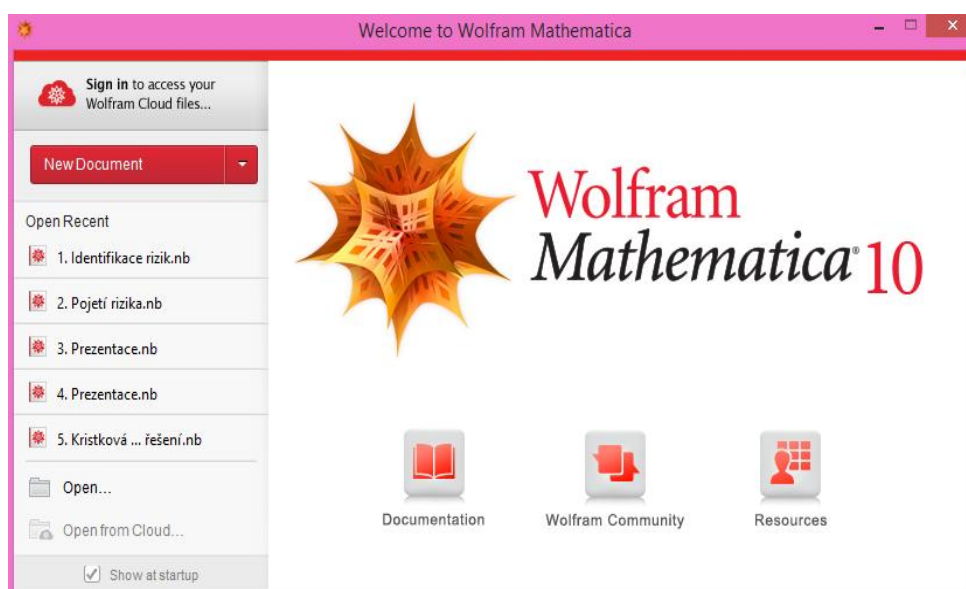
³FRIEDRICH, V., *Mathematica na počítači pro nematematiky*, s. 5.

⁴Srov. FŇUKAL, M., SMRČKA, D., a KLADIVO P., *Software Mathematica pro geography*, s. 5.

⁵Srov. CHRAMCOV, B., *Základy práce v prostředí Mathematica*, s. 11.

1.5 Prostředí

Po spuštění programu se zobrazí uvítací okno (obrázek 1), které vám nabídne několik možností, jak dále pokračovat. Nejdůležitější pro uživatele je sloupec vlevo. V tomto sloupci si můžeme vybrat, jestli chcete vytvořit nový soubor (záložka „New Document“) nebo otevřít starý (již vytvořený) dokument. Nabídka nedávno vytvořených souborů, které můžeme v programu otevřít, je vypsaná pod příkazem „Open Recent“. U spodního okraje jsou tři ikony, které mají doplňkový charakter. Pod ikonou „Documentation“ se skrývá kompletní nápověda k programu, po kliknutí na zbylé dvě ikony se otevrou webové stránky www.wolfram.com. Na bílém pozadí vidíme název a verzi programu (zde verze č. 10).

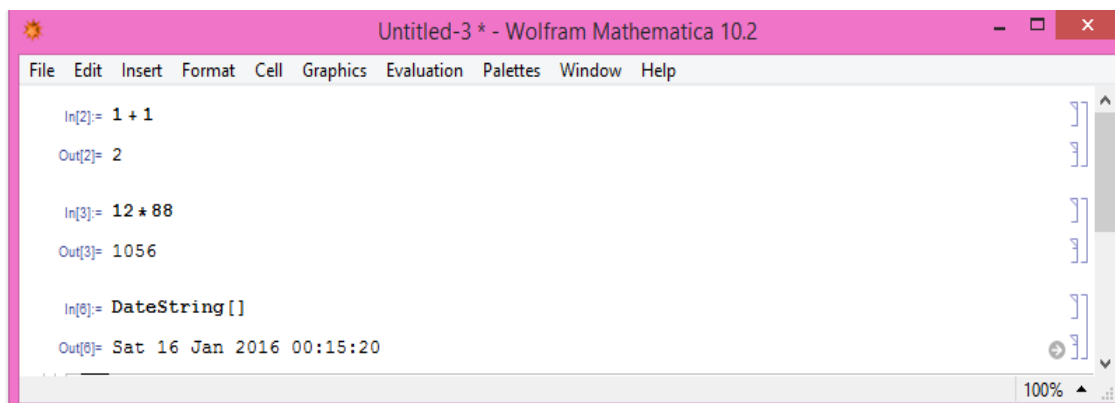


Obrázek 1 Uvítací okno programu

zdroj:vlastní

Po rozkliknutí záložky nový dokument, se otevře nabídka, ze které si vyberete, jaký typ nového dokumentu chcete vytvořit. V nabídce je typ „Notebook“, „Slide Show“, „Demonstration“ a několik dalších. V této bakalářské práci budeme nejvíce využívat prostředí zvané „Notebook“. Jak už označení napovídá, jedná se v podstatě o prázdné okno, do kterého zadáváme příkazy.

Jak vypadá prostředí Notebook, vidíme na obrázku 2. Notebook je hlavní rozhraní, přes které uživatel komunikuje s programem. Pro lepší seznámení s programem si notebook popíšeme. Jak můžeme vidět, nahoře se nachází lišta s různými nástoji a pomocníky pro práci s software Mathematica. Na liště je celkem deset záložek, každá záložka po rozkliknutí nabízí další ovládací prvky pro program. V následujícím odstavci si jednotlivé záložky stručně představíme.



Obrázek 2 Notebook

zdroj:vlastní

První zleva je záložka „File“. Najdeme na ní funkce jako uložit dokument, vytisknout, otevřít jiný dokument, vytvořit nový, zavřít. Další je „Edit“ neboli úprava dokumentu. Pomocí této záložky můžeme vrátit poslední krok zpět, kopírovat, vložit, najít a další. Hned vedle se nachází záložka „Insert“, která umožňuje přímé vložení různých prvků, např. obrázek, předešlý výpočet či výsledek, matice. Za ní je záložka „Format“. Pomocí této záložky si uživatel může nastavit vzhled samotného notebooku, barvu písma, pozadí, typ vstupních údajů, atd. Následuje záložka „Cell“. Touto záložkou ovládá uživatel nastavení buněk. Co je to buňka si vysvětlíme v následujícím odstavci. Další je záložka „Graphics“. Pomocí této záložky můžeme do notebooku nakreslit nějaký obrázek, podobně jako v programu Malování. Sedmá záložka se jmenuje „Evaluation“ a slouží k aktualizaci dat v Notebooku. Vedle je záložka „Palettes“, která usnadňuje práci s programem prostřednictvím tzv. asistentů. Předposlední záložka se nazývá „Window“ a umožňuje nastavení okna notebooku jako celku, např. velikost okna, přidání dalších listů, zobrazení na celou obrazovku. Poslední záložka je „Help“. V této záložce se nachází kompletní nápověda k programu a technické informace o programu.

Samotný notebook je rozdělený na buňky. Ty jsou znázorněny na obrázku 2 pomocí hranatých závorek při pravém okraji dokumentu. Začátek buňky značí vodorovný kurzor. Do buňky můžeme napsat různé typy informací (číselné i textové), to jak je bude program interpretovat záleží na nastavení stylu buňky.

Program nabízí různé styly, nejpoužívanější styly (vidíme je i na obrázku 2) jsou „Input“ a „Output“, zkratka In a Out. Tyto styly používáme v případě, když chceme od programu nějaký výsledek, nějaký výstup. Do buňky Input zadáme požadavek (na obrázku 2 chceme vypočítat příklad 1+1) a v buňce Output se zobrazí výsledek (na

obrázku vidíme číslo 2). Aby se nám výsledek zobrazil, musíme buňku tzv. uzavřít. To uděláme stisknutím klávesy ENTER na numerické klávesnici nebo ENTER+SHIFT. Dalším hodně používaným stylem je Text. Do takto nastavené buňky můžeme psát cokoliv, věty, poznámky, samostatná slova a upravovat to jako u klasického textového editoru (barva písma, velikost, styl písma, zarovnání). Díky tomuto nastavení můžeme psát v programu Mathematica souvislou práci, různé poznámky a popisy. Textovou buňku uzavřeme pomocí vodorovného kurzoru, tak že ho umístíme pod buňku a stiskneme levé tlačítko na myši. Tímto krokem se přemístíme do nové buňky a můžeme pokračovat v práci.

Buňky lze libovolně upravovat, můžeme je přesouvat po celém notebooku, kopírovat je, nastavit formát (písmo, pozadí, zarovnání písma), vymazat celou buňku, aktualizovat buňku. Můžeme označit více buňek najednou a formátovat je podobně jako buňku samostatnou. Pokud klikneme na hranatou závorku uzavírající buňku pravím tlačítkem myši, otevře se celá nabídka formátování buňek. Pokud chceme pracovat s jednotlivými buňkami, musíme vždy levým tlačítkem na myši kliknout na hranatou závorku příslušné buňky, tím tuto buňku vybereme.

Notebook není určen jen pro výpočty, dá se upravovat jako klasický textový dokument, můžeme v něm tvořit tabulky, či jiné grafické výstupy. S notebookem je možné pracovat jako jiným textovým editorem, můžeme ho ukládat, kopírovat, smazat, vytisknout, nebo přepisovat. Slouží také k vytváření prezentací. Notebook lze uložit i jako formáty PDF, TXT, HTML.

1.6 Programovací jazyk

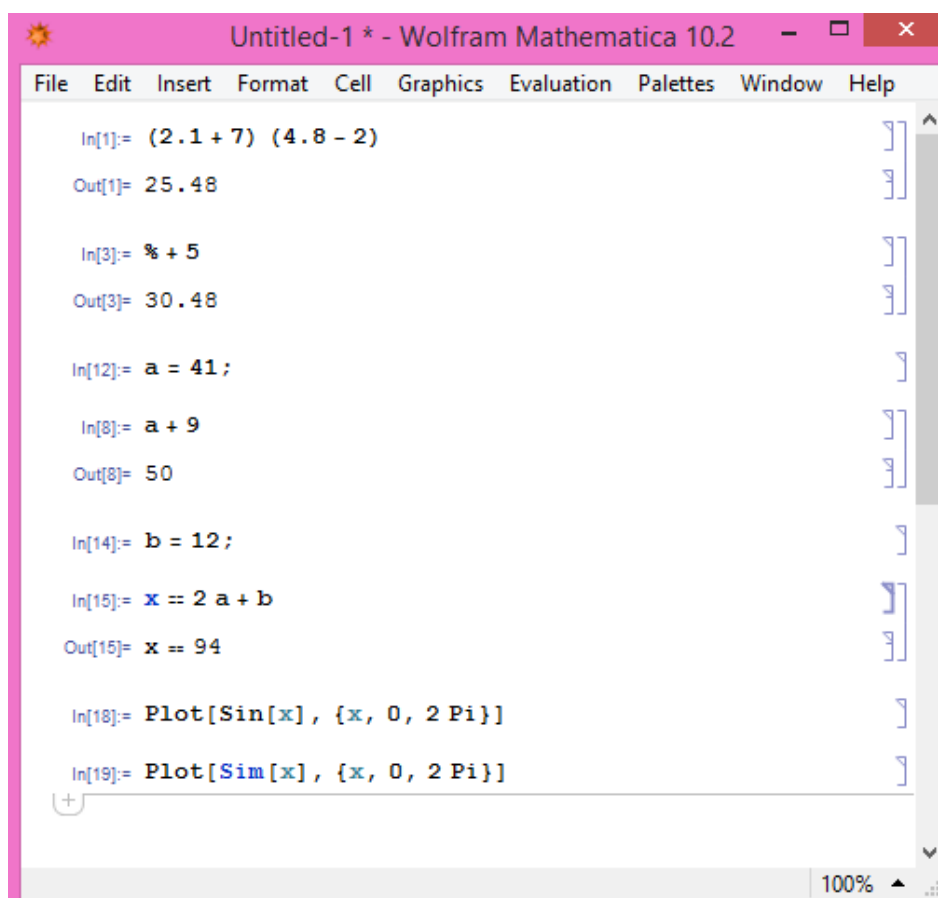
Jak bylo v předešlé podkapitole řečeno, notebook dělíme na jednotlivé buňky. Pokud od programu Mathematica požadujeme jakýkoliv výstup (prostřednictvím buňky označené Out), musíme správně napsat svůj požadavek do vstupní buňky In. K tomu slouží několik pravidel, která nám zajistí, že program pochopí náš požadavek a vrátí nám správný výsledek.

Program rozlišuje několik druhů závorek. Kulaté závorky slouží k seskupování a klasicky při aritmetických výpočtech, např. $(8-4)*(12+4)$. Hranaté závorky slouží při zápisu funkce, např. $\sin[x]$. Do složených závorek se zapisují seznamy, výpisy, intervaly, např. $\{1,12,28,32\}$. A poslední dvojité hranaté závorky slouží pro indexování, např. $m[[3]]$.

Mezi další pravidla patří:

- název funkce musíme psát s velkým počátečním písmenem;
- při zápisu desetinného čísla se používá tečka;
- středník za příkazem slouží k tomu, že se příkaz provede, ale v notebooku se výsledek nezobrazí;
- názvy proměnných a funkcí, které jsme si definovali sami, píšeme s malými písmeny;
- pokud je název funkce modrý, je napsán špatně. Až název funkce napíšeme správně, barva se změní na černou.
- při násobení můžeme používat místo symbolu krát, mezeru;
- rovnost v rovnici musíme zapsat jako „==“;
- symbol „%“ se používá, když chceme definovat novou proměnou;
- symbol „%“ se používá, když chceme použít předešlý výsledek.

Příklady použití jednotlivých pravidel vidíme níže na obrázku 3:



```
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

In[1]:= (2.1 + 7) (4.8 - 2)
Out[1]= 25.48

In[3]:= % + 5
Out[3]= 30.48

In[12]:= a = 41;

In[8]:= a + 9
Out[8]= 50

In[14]:= b = 12;

In[15]:= x == 2 a + b
Out[15]= x == 94

In[18]:= Plot[Sin[x], {x, 0, 2 Pi}]

In[19]:= Plot[Sim[x], {x, 0, 2 Pi}]

+
100%
```

Obrázek 3 Příklady použití pravidel

zdroj: vlastní

2 ÚVOD DO STATISTIKY

V této kapitole se budeme zabývat pojmem statistika a jeho dalšími významy. Dále tato kapitola bude obsahovat krátkou historii statistiky a nějaké základní pojmy, se kterými se můžeme ve statistice setkat.

2.1 Pojem statistika

Pojem statistika má v současné době několik významů. Nejprve si pod tímto pojmem můžeme představit nějaká konkrétní data či číselné údaje, které jsme získali např. z výzkumu, z dotazování. Dále pojem statistika představuje reálnou činnost, kdy sbíráme, zpracováváme a vyhodnocujeme údaje o hromadných jevech a nakonec statistika je „teoretická disciplína, která se zabývá metodami, sloužícími k popisu odhalování zákonitostí při působení podstatných, relativně stálých činitelů na hromadné jevy.“⁶ V této bakalářské práci se budeme věnovat statistice jako teoretické disciplíně.

2.2 Historie a současnost

První zmínky o statistice se objevují už ve starověkých říších. V této době se nejvíce sestavovaly různé seznamy s daty o obyvatelstvu, které se využívaly nejčastěji pro vyměřování daní. První statistické analýzy tehdejších událostí se začaly vytvářet v 17. století a jsou spojeny se jmény John Graunt (1620-1674) a William Petty (1623-1687). V polovině 18. století Gottfried Achenwall (1719-1772) vymyslel slovo statistika a vznikl tak nový vědní obor. Další rozvoj zažila statistika v 19. a 20. století, tomuto rozvoji napomohly i základy teorie pravděpodobnosti, od vědců Jacob Bernoulli, Karel F. Gauss a mnoho dalších. Další rozvoj nastal ve 20. stoletím a v 70. letech a to díky rychlému pokroku v oblasti výpočetní techniky a vzniku statistického výpočetního prostředí.⁷

V dnešní době se bez statistiky neobejdeme. Každý vědní obor, ať je to medicína, fyzika, biologie či další přírodní i technické obory, pracuje s hromadnými daty či číselnými údaji a používá při práci statistické metody. Statistiku nalezneme i v hospodářské oblasti nebo při analýze sociálněekonomických jevů. Pro potřeby hospodářské oblasti vznikly dokonce speciální disciplíny a části statistiky.⁸

⁶ HINDLS, R., HRONOVÁ, S., SEGER, J., *Statistika pro ekonomy*, s. 12.

⁷ Srov. tamtéž.

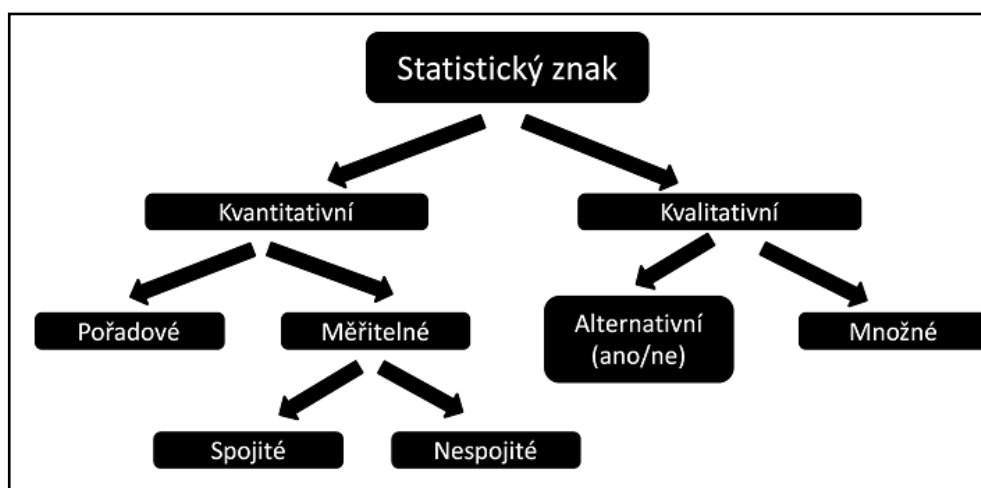
⁸ Srov. tamtéž.

2.3 Základní pojmy

V kapitole Základní pojmy se budeme zabývat vybranými základními pojmy ze statistiky. Ve statistice se můžeme setkat s nespočtem různých pojmů, pro účely této práce budou objasněny pouze ty pojmy, které se vyskytují v dalších kapitolách.

Statistická jednotka: „Prvek souboru, u něhož jsou sledovány různé vlastnosti. Statistickými jednotkami mohou být například osoby, firmy, města, automobily, zvířata.“⁹

Statistický znak: konkrétní vlastnost statistické jednotky (např. věk). Pro tento pojem se také užívá výraz statistická proměnná. Statistický znak dosahuje pro jednotlivé statistické jednotky určitých hodnot, které jsou zpravidla rozdílné (např. všechny osoby nemají přesně stejnou váhu). Statistické proměnné dělíme na různé kategorie podle obrázku 4. Kvantitativní znaky dosahují číselné hodnoty (různé hodnoty váhy), naopak kvalitativní jsou vyjádřeny slovně (forma studia). Kvalitativní dále dělíme na alternativní (proměnné nabývají pouze dvou hodnot, studuje-nestuduje) a množné (proměnné nabývají více než dvou hodnot, nejvyšší dosažené vzdělání). Kvantitativní znaky dělíme na pořadové (můžeme je seřadit od nejnižšího po nejvyšší, hodnocení ve škole) a měřitelné (hodnoty můžeme mezi sebou porovnávat, počet pracovníků jednotlivých oddělení). Měřitelné znaky se dále dělí na nespojité (znaky mají charakter celočíselných hodnot, počet obyvatel), a spojité (nabývají libovolná čísla, výše mzdy).¹⁰



Obrázek 4 Rozdělení statistických znaků

zdroj: http://kantaylis.com/zakladni_statisticke_pojmy.html.

⁹ Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 7.

¹⁰ Srov. tamtéž s. 8.

Statistický soubor: skupina, do které jsme zahrnuli všechny statistické jednotky, u nichž sledujeme totožné statistické znaky. Můžeme se setkat s dvěma typy statistických souborů a to základní soubor (např.: obyvatelstvo ČR) a výběrový soubor (vybereme pouze zástupce ze základního souboru).¹¹

Statistický soubor s jedním argumentem: soubor se všemi statistickými jednotkami, které mají pouze jeden statistický znak X. Např. u skupiny dětí ve věku 10 až 15 let sledujeme jejich váhu.

Statistický soubor se dvěma argumenty: soubor, u něhož sledujeme statistické jednotky, kdy každá jednotka má dva statistické znaky X a Y. Tyto znaky mohou být spojité či diskrétní. Při sledování se zajímáme o hodnoty obou znaků, zkoumáme je jak samostatně tak i ve vzájemném vztahu. Např. při sledování souboru lidí se můžeme zajímat o dva znaky-výška a váha.¹²

Rozsah souboru: počet všech prvků ve statistickém souboru. Rozsah souboru má značku písmeno N. Např. počet studentů ve skupině, u které zjišťujeme výšku studentů.¹³

¹¹ Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 7.

¹² Srov. OPITKA, P., ŠMAJSTRLA, V., *Pravděpodobnost a statistika*, s. 165

¹³ Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 7.

3 CHARAKTERISTIKY STATISTICKÉHO SOUBORU

V další kapitole této bakalářské práce zůstaneme u tématu statistika, konkrétně se budeme věnovat problematice charakteristiky statistického souboru. Někdy se můžeme setkat s přívlastkem empirické vlastnosti statistického souboru. Tento přívlastek znamená, že jsme hodnoty pro výpočet charakteristik získali ze statistického šetření. Ve statistice se často setkáváme s velkým objemem dat či informací, a bylo by obtížné pracovat s nimi najednou. Proto používáme určité charakteristiky, které nám pomohou popsat statistický soubor jediným číselným údajem.

Ve statistice rozlišujeme charakteristiky polohy, variability, šikmosti, špičatosti. První dvě charakteristiky jsou velmi používané, zbylé dvě jsou při práci se statistickým souborem oblíbeny méně. Každá charakteristika zkoumá statistický soubor z jiného úhlu pohledu, pouze tak získáváme velmi komplexní informace.

V této bakalářské práci se budeme postupně zabývat všemi charakteristikami, představíme si jednotlivé zástupce od každé charakteristiky a uvedeme konkrétní příklady, které budou řešeny v programu Mathematica.

3.1 Charakteristiky polohy

Charakteristiky polohy nám sdělují, kde na číselné ose jsou data rozloženy a zároveň se snaží určit hodnotu neboli „střed“ kolem které se data pohybují. Charakteristiky dělíme na dvě kategorie:

- do první skupiny řadíme všechny druhy průměrů (aritmetický, geometrický, harmonický) a velikost těchto charakteristik závisí na velikosti všech hodnot znaku;
- ve druhé skupině se nachází modus a medián a velikost těchto dvou charakteristik nezávisí na velikosti všech hodnot znaku.¹⁴

¹⁴ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 99

3.1.1 Aritmetický průměr

Máme vybraný statistický soubor, který má znak X a tento znak nabývá hodnot x_1, x_2, \dots, x_n (hodnoty nejsou uspořádány do tabulky četností a neopakují se). Potom můžeme zjistit tzv. prostý aritmetický průměr (značka \bar{x}), který má vzorec:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

kde n je celkový počet hodnot. Výsledek se udává ve stejných měrných jednotkách, ve kterých jsou uvedeny jednotlivé hodnoty proměnné (kg, m, ks). Pokud jsou zjištěné hodnoty $x_1 \dots x_n$ uspořádány do tabulky četností a některé hodnoty se opakují, počítáme tzv. vážený aritmetický průměr podle vzorce:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i}, \quad (2)$$

kde n představuje četnost, neboli kolikrát se stejné hodnoty znaku opakují. Tyto dva vzorce jsou nejpoužívanější a dají se uplatnit ve všech úlohách statistiky.¹⁵

Aritmetický průměr má několik vlastností, některé využíváme při výpočtech a jiné mají teoretický význam:

- aritmetický průměr konstanty c je roven této konstantě $\bar{c} = c$;
- pokud přičteme ke každé hodnotě znaku konstantu, vzroste o tuto konstantu i aritmetický průměr $\bar{x} + k = \overline{x + k}$;
- pokud jednotlivé hodnoty znaku vynásobíme konstantou, průměr je touto konstantou také násoben $\bar{x} \times c = \overline{x \times c}$;
- násobíme-li jednotlivé četnosti konstantou, aritmetický průměr se nezmění;
- součet odchylek hodnot proměnné a jejího aritmetického průměru je nulový;
- extrémní hodnoty znaku mají vliv na aritmetický průměr.¹⁶

¹⁵ Srov. HINDLS, R., HRONOVÁ, S., SEGER, J., *Statistika pro ekonomy*, s. 30

¹⁶ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 101

Příklad č. 1

Počet dopravních nehod na území České republiky v letech 1997 –2014 vidíme v tabulce 1. Určete prostý aritmetický průměr.

Tabulka 1 Počet nehod v letech

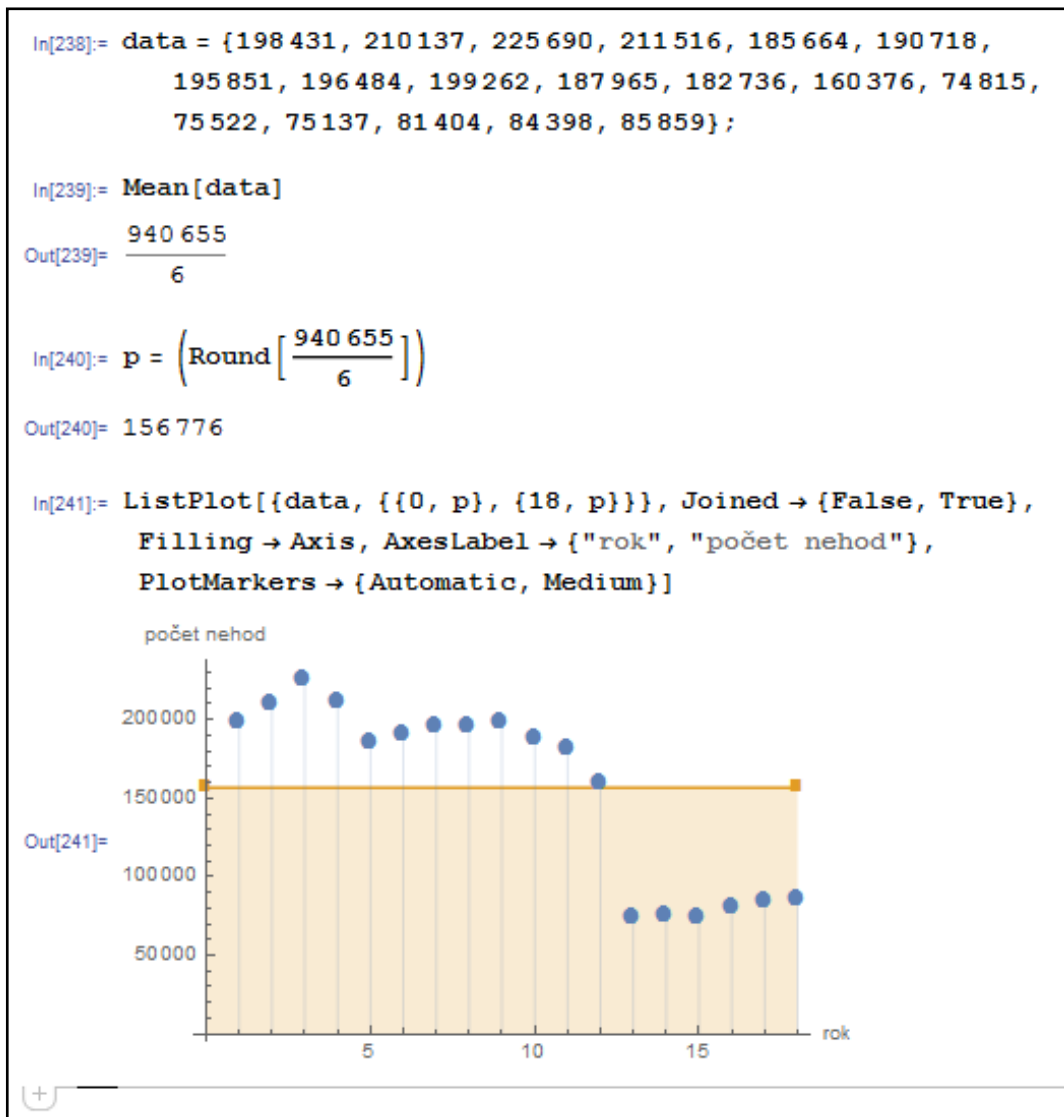
Rok	Počet nehod	Rok	Počet nehod
1997	198 431	2006	187 965
1998	210 137	2007	182 736
1999	225 690	2008	160 376
2000	211 516	2009	74 815
2001	185 664	2010	75 522
2002	190 718	2011	75 137
2003	195 851	2012	81 404
2004	196 484	2013	84 398
2005	199 262	2014	85 859

zdroj: Český statistický úřad, www.czso.cz/csu/czso/ceska-republika-od-roku-1989-v-cislech

Řešení:

Pro řešení použijeme software Mathematica (obrázek 5), který má pro prostý aritmetický průměr předefinovanou funkci *Mean*. Nejprve jsme si zjištěná čísla o počtu nehod vložili do programu, vytvořili jsme si seznam s názvem *data*. Následně jsme použili funkci *Mean*, do hranatých závorek jsme vložili předpřipravený seznam hodnot (argument funkce). Výsledek příkladu vidíme v buňce *Out[239]*. Tento výsledek je zlomek, který nemá velkou vypovídající schopnost. Proto jsme použili funkci *Round*, která nám zlomek vyjádřila v celých číslech. Závěr: v období od r. 1997 do r. 2014 se stalo průměrně 156 776 dopravních nehod každý rok.

Program Mathematica je velice vhodný pro tvoření grafů a jiných grafických výstupů. Na obrázku 5 vidíme graf, který vyjadřuje kolísání zjištěných hodnot kolem aritmetického průměru. Modré body značí počet nehod v jednotlivých letech, oranžová přímka vyjadřuje námi vypočítaný průměr. Pro vykreslení grafu jsme použili funkci *ListPlot*. Pro označení jednotlivých os u grafu jsme použili funkci *AxesLabel*. Pro nastavení formátu modrých bodů jsme použili funkci *PlotMarkers*. K zvýraznění prostoru pod křivkami slouží příkaz *Filling*.



Obrázek 5 Řešení příkladu č. 1

zdroj: vlastní

Příklad č. 2

Firma z důvodu úspory nákladů prováděla průzkum. Byly vybrány statistické jednotky - pracovníci, u kterých se zjišťoval statistický znak tarifní třída (kvantitativní znak). Vypočtete průměrnou tarifní třídu pracovníků. Další údaje potřebné pro výpočet vidíme na obrázku 6, vytvořená tabulka v Out[7].¹⁷

Řešení:

Výpočet vidíme na obrázku 6 (viz následující strana). Nejprve jsme si definovali seznamy dat a označili jsme je *ni* a *xi*. Protože z těchto seznamů chceme vytvořit tabulku, použili jsme příkaz *Partition*, který nám spáruje vždy 2 hodnoty. Z takto vybraných hodnot vytvoříme pomocí příkazu *TableForm* tabulku. K tabulce přidáme pomocí příkazu *Prepend* záhlaví a pomocí příkazu *Append* poslední řádek tabulky, ve kterém máme součet pracovníků. Výslednou tabulku vidíme v Out[7].

Na výpočet použijeme vzorec pro vážený aritmetický průměr. Tento vzorec není v programu Mathematica předdefinovaný, proto jsme si definovali novou funkci VAP, která se rovná vzorci pro vážený aritmetický průměr a do tohoto vzorce jsme dosadili hodnoty *ni* a *xi*. V Out[12] vidíme námi zadaný vzorec, ke kterému jsme rovnou přidali příkaz //N pro zobrazení výsledku ve formě numerického čísla. Pomocí funkce *Round* jsme tento výsledek zaokrouhlili. Závěr: průměrná tarifní třída mezi 75 zaměstnanci je 6. tarifní třída.

¹⁷ Srov. HINDLS, R., HRONOVÁ, S., SEGER, J., *Statistika pro ekonomy*, s. 30

```

In[1]:= xi = {3, 4, 5, 6, 7};

In[2]:= ni = {2, 11, 19, 27, 16};

In[3]:= hodnoty = Partition[Riffle[xi, ni], 2]
Out[3]= {{3, 2}, {4, 11}, {5, 19}, {6, 27}, {7, 16}}

In[4]:= tab1 = (hodnoty // TableForm)
Out[4]/TableForm=
  3    2
  4   11
  5   19
  6   27
  7   16

In[5]:= tab3 =
  (Prepend[hodnoty, {"Tarifní třída xi", "Počet pracovníků ni"}])
Out[5]= {{Tarifní třída xi, Počet pracovníků ni},
  {3, 2}, {4, 11}, {5, 19}, {6, 27}, {7, 16}}

In[6]:= t1 = (Append[tab3, {"celkem", Total[ni]}])
Out[6]= {{Tarifní třída xi, Počet pracovníků ni}, {3, 2},
  {4, 11}, {5, 19}, {6, 27}, {7, 16}, {celkem, 75}}

In[7]:= Grid[t1, Frame -> All, Background -> {{None}, {LightBlue}}]
Out[7]=


| Tarifní třída xi | Počet pracovníků ni |
|------------------|---------------------|
| 3                | 2                   |
| 4                | 11                  |
| 5                | 19                  |
| 6                | 27                  |
| 7                | 16                  |
| celkem           | 75                  |



In[12]:= VAP[x1_, x2_, x3_, x4_, x5_, n1_, n2_, n3_, n4_, n5_] =
  (x1 * n1 + x2 * n2 + x3 * n3 + x4 * n4 + x5 * n5) / (n1 + n2 + n3 + n4 + n5)
  VAP[3, 4, 5, 6, 7, 2, 11, 19, 27, 16] // N
Out[12]= 
$$\frac{n1 x1 + n2 x2 + n3 x3 + n4 x4 + n5 x5}{n1 + n2 + n3 + n4 + n5}$$


Out[13]= 5.58667

In[14]:= Round[%]
Out[14]= 6

```

Obrázek 6 Řešení příkladu č. 2

zdroj: vlastní

3.1.2 Geometrický průměr

Geometrický průměr používáme v případě, že chceme vypočítat průměrný koeficient růstu, pracujeme zde s určitými přírůstky k dané hodnotě. Geometrický průměr členíme, tak jako aritmetický, na prostý geometrický průměr a vážený geometrický průměr. Nejprve si vysvětlíme prostý geometrický průměr. Uvažujeme n kladných hodnot proměnné X , hodnoty nejsou uspořádány do tabulky četností a jsou kladné, potom vzorec pro výpočet má tvar:

$$\bar{x}_G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}, \quad (3)$$

kde n značí počet hodnot a x_1, x_2, \dots, x_n jsou zjištěné hodnoty, ze kterých chceme geometrický průměr vypočítat (např. koeficienty růstu, přírůstky). U váženého geometrického průměru uvažujeme řadu n hodnot proměnné X (hodnoty jsou uspořádány do tabulky četností) a zároveň uvažujeme k variant této proměnné, potom použijeme vzorec:

$$\bar{x}_G = \sqrt[n]{x_1^{n_1} \times x_2^{n_2} \times \dots \times x_k^{n_k}}, \quad (4)$$

kde n představuje počet hodnot, x_1, x_2, \dots, x_k jsou zjištěné hodnoty a n_1, n_2, \dots, n_k vyjadřují počet opakování stejné hodnoty. Oba dva typy geometrického průměru používáme v běžné ekonomické praxi minimálně. Vlastnosti geometrického průměru jsou analogické, jako vlastnosti průměru aritmetického, které jsme uvedly v kapitole 3.1.1 Aritmetický průměr.¹⁸

¹⁸ Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 24.

Příklad č. 3

Určete průměrný koeficient růstu počtu výběrového statistického souboru (obyvatelé v Praze) v letech 2001–2009. Údaje o počtu obyvatel za jednotlivé roky vidíme v tabulce 2.

Tabulka 2 Počet obyvatel v Praze

Rok	Počet obyvatel
2001	1 160 118
2002	1 161 938
2003	1 165 581
2004	1 170 571
2005	1 181 610
2006	1 188 126
2007	1 212 097
2008	1 233 211
2009	1 249 026

zdroj: ČSU, www.czso.cz/csu/czso/databaze-demografickych-udaju-za-obce-cr

Řešení:

Řešení příkladu v software Mathematica vidíme na obrázku 7. Nejprve jsme si vypočítali koeficienty růstu v jednotlivých letech a tyto koeficienty jsme označili k1-k8. Následně jsme využili předefinovanou funkci *GeometricMean*, abychom vypočítali prostý geometrický průměr. Výsledek ve tvaru zlomku vidíme v Out[61]. Protože zlomek nemá dostatečně velkou vypovídající hodnotu, převedli jsme si ho na desetinné číslo. Závěr: průměrný roční koeficient růstu počtu obyvatel je 0,927 %.

```
In[59]:= k1 = 1161938 / 1160118 ; k2 = 1165581 / 1161938 ; k3 = 1170571 / 1165581 ;
k4 = 1181610 / 1170571 ; k5 = 1188126 / 1181610 ; k6 = 1212097 / 1188126 ;
k7 = 1233211 / 1212097 ;
k8 = 1249026 / 1233211 ;

In[61]:= GeometricMean[{k1, k2, k3, k4, k5, k6, k7, k8}]
Out[61]=  $\left(\frac{208171}{193353}\right)^{1/8}$ 

% // N
Out[62]= 1.00927
```

Obrázek 7 Řešení příkladu č. 3

zdroj: vlastní

Příklad č. 4

Vypočítejte průměrný přírůstek (úbytek) základního statistického souboru (obyvatelstvo v České republice) v roce 2015. Údaje o jednotlivých statistických jednotkách (kraj) zahrnutých do statistického souboru vidíme v tabulce 3.

Tabulka 3 Přírůstek (úbytek) v krajích v roce 2015

Kraj	Přírůstek (úbytek) obyvatel v %	Kraj	Přírůstek (úbytek) obyvatel v %
Praha	1,9	Královéhradecký	-0,5
Středočeský	1,2	Pardubický	-0,3
Liberecký	0,3	Vysočina	0,3
Ústecký	-1,1	Olomoucký	-0,8
Karlovarský	-1,8	Jihomoravský	0,6
Plzeňský	-0,8	Moravskoslezský	-1,4
Jihočeský	-0,5	Zlínský	-0,8

zdroj: Český statistický úřad,

<https://www.czso.cz/csu/xm/prirustek-ubytek-obyvatelstva-prirozenou-menou-v-krajich-cr-v-roce-201x>

Řešení:

Na začátku řešení příkladu jsme si z dokumentu Microsoft Excel importovali data se základními údaji. Následně si musíme přírůstky (úbytky), které jsou v tabulce v procentech, vyjádřit jako desetinné číslo. Určíme si, že rok 2014 je základna (tzn. 100%) a pokud v roce 2015 vzroste počet obyvatel např. o 1,9%, dostaneme se na 101,9% minulého roku. Toto číslo vydělíme 100 a dostaneme desetinné číslo 1,019. Pokud počet obyvatel klesl oproti minulému roku, musíme odečítat, např. úbytek 0,8% je 99,2% minulého roku, převedeno na desetinné číslo 0,992. Všechny přepočty vidíme na obrázku 8 v Out[62]. K takto doplněným údajům jsme dále přidali záhlaví tabulky a do prostředního sloupce jsme přidali symbol procenta. Výslednou tabulku vidíme na obrázku 9 v Out[65].

Z tabulky jsme pomocí příkazu *Range* vybrali jen poslední sloupec, se kterým budeme pracovat dále. Z vybraných dat vypočítáme pomocí funkce *GeometricMean* geometrický průměr. Závěr: průměrný úbytek obyvatel v krajích v roce 2015 byl přibližně 0,27 %.


```

In[81]= tabulka1 = Import["F:\\přklady\\Příklad 4_mezivýpočty.xlsx"][[1]]
Out[81]= {{Praha, 1.9}, {Středočeský, 1.2}, {Liberecký, 0.3}, {Ústecký, -1.1},
          {Karlovarský, -1.8}, {Plzeňský, -0.8}, {Jihočeský, -0.5},
          {Královéhradecký, -0.5}, {Pardubický, -0.3}, {Vysočina, 0.3}, {Olomoucký, -0.8},
          {Jihomoravský, 0.6}, {Moravskoslezský, -1.4}, {Zlínský, -0.8}}

In[82]= přečet = tabulka1 /. {a_, b_} -> {a, b, (b+100)/100}
Out[82]= {{Praha, 1.9, 1.019}, {Středočeský, 1.2, 1.012}, {Liberecký, 0.3, 1.003},
          {Ústecký, -1.1, 0.989}, {Karlovarský, -1.8, 0.982}, {Plzeňský, -0.8, 0.992},
          {Jihočeský, -0.5, 0.995}, {Královéhradecký, -0.5, 0.995},
          {Pardubický, -0.3, 0.997}, {Vysočina, 0.3, 1.003},
          {Olomoucký, -0.8, 0.992}, {Jihomoravský, 0.6, 1.006},
          {Moravskoslezský, -1.4, 0.986}, {Zlínský, -0.8, 0.992}}

In[83]= tabulka2 = Prepend[přečet,
                          {"Kraj", "Přírůstek (úbytek) obyvatel v%", "Přečet na desetinné číslo"}];
In[84]= tabulka2 = tabulka2 /. {a_, b_, c_} -> {a, ToString[b] <> "%", c}
Out[84]= {{Kraj, Přírůstek (úbytek) obyvatel v%, Přečet na desetinné číslo},
          {Praha, 1.9%, 1.019}, {Středočeský, 1.2%, 1.012}, {Liberecký, 0.3%, 1.003},
          {Ústecký, -1.1%, 0.989}, {Karlovarský, -1.8%, 0.982}, {Plzeňský, -0.8%, 0.992},
          {Jihočeský, -0.5%, 0.995}, {Královéhradecký, -0.5%, 0.995},
          {Pardubický, -0.3%, 0.997}, {Vysočina, 0.3%, 1.003},
          {Olomoucký, -0.8%, 0.992}, {Jihomoravský, 0.6%, 1.006},
          {Moravskoslezský, -1.4%, 0.986}, {Zlínský, -0.8%, 0.992}}

```

Obrázek 9 Přečet na desetinná čísla

zdroj:vlastní

```

In[85]= Grid[tabulka2, Frame -> All, Background -> {{LightBlue}, {LightBlue}},
            Spacings -> {0.5, 0.8}]

```

Kraj	Přírůstek (úbytek) obyvatel v%	Přečet na desetinné číslo
Praha	1.9%	1.019
Středočeský	1.2%	1.012
Liberecký	0.3%	1.003
Ústecký	-1.1%	0.989
Karlovarský	-1.8%	0.982
Plzeňský	-0.8%	0.992
Jihočeský	-0.5%	0.995
Královéhradecký	-0.5%	0.995
Pardubický	-0.3%	0.997
Vysočina	0.3%	1.003
Olomoucký	-0.8%	0.992
Jihomoravský	0.6%	1.006
Moravskoslezský	-1.4%	0.986
Zlínský	-0.8%	0.992

```

In[86]= data = Flatten[tabulka2[[Range[2, 15], Range[3, 3]]]]
Out[86]= {1.019, 1.012, 1.003, 0.989, 0.982, 0.992,
          0.995, 0.995, 0.997, 1.003, 0.992, 1.006, 0.986, 0.992}

In[87]= výsledek = GeometricMean[data]
Out[87]= 0.997309

In[88]= (1 - výsledek) * 100
Out[88]= 0.269112

```

Obrázek 8 Řešení příkladu č. 4

zdroj: vlastní

3.1.3 Harmonický průměr

Harmonický průměr využíváme v praxi tam, kde chceme zjistit například průměrnou rychlost nebo průměrnou dobu pro splnění určitého úkonu (všechny jednotky pracují na úkolu současně, společná práce). Obecně používáme harmonický průměr tam, kde má smysl se zabývat převrácenými hodnotami. Pro výpočet harmonického průměru potřebujeme pouze kladné hodnoty. Harmonický průměr se dělí na prostý a vážený. Prostý harmonický průměr vypočítáme podle vzorce:

$$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}, \quad (5)$$

n znamená počet hodnot proměnné X a x_i jsou jednotlivé zjištěné hodnoty. Hodnoty nejsou uspořádány do tabulky četností a neopakují se. Pokud jsou zjištěné hodnoty uspořádány do tabulky četností a některá data se opakují, používáme pro výpočet vzorec pro vážený harmonický průměr:

$$\bar{x}_H = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k \frac{n_i}{x_i}}, \quad (6)$$

x_i představují jednotlivé zjištěné hodnoty znaku X a n_i jsou váhy jednotlivých hodnot znaku X . Harmonický průměr je vždy menší než geometrický průměr i menší než aritmetický průměr.¹⁹

¹⁹ Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 25.

Příklad č. 5

Firma chce optimalizovat svou výrobu, proto chce zjistit průměrnou dobu potřebnou k vykonání požadované práce. Vedoucí zaměstnanci firmy stanovili statistický soubor s jedním argumentem, který obsahuje pět zaměstnanců, u kterých zjišťujeme, za jak dlouho splní požadovanou práci. První zvládne svou práci udělat za 5 minut, druhý to stihne za 8 minut, třetí za 7 minut, čtvrtý za 4 minuty a pátý za 6 minut. Určete průměrnou dobu, která je potřebná k vykonání požadované práce.²⁰

Řešení:

Pro řešení tohoto příkladu (řešení vidíme na obrázku 10) použijeme vzorec pro prostý aritmetický průměr, tento vzorec nalezneme v programu Mathematica pod funkcí *HarmonicMean*. Jednotlivé zjištěné hodnoty, v tomto případě počet minut, musíme napsat do složených závorek (určených pro seznam) a ještě do hranatých závorek (argument funkce). Výsledek se zobrazí ve tvaru zlomku, proto je dobré tento zlomek převést do tvaru desetinného čísla. Protože chceme zjistit průměrnou dobu, musíme si toto desetinné číslo vyjádřit ve formátu hh:mm:ss. K tomu slouží funkce *DateString*. Konečný výsledek vidíme v Out[3]. Závěr: k vykonání zadané práce potřebujeme průměrně 5 minut a 39 sekund.

```
In[1]:= HarmonicMean[{5, 8, 7, 4, 6}]
Out[1]=  $\frac{4200}{743}$ 

In[2]:= % // N
Out[2]= 5.65276

In[3]:= DateString[% * 60, "Time"]
Out[3]= 00:05:39
```

Obrázek 10 Řešení příkladu č. 5

zdroj: vlastní

²⁰ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 104

Příklad č. 6

Firma vyrábějící ovocné marmelády nakupuje od dodavatelů čerstvé ovoce. Při prvním nákupu stálo ovoce 40 Kč/Kg, firma nakoupila za 50 Kč. Při druhém nákupu stálo ovoce 50 Kč/Kg a firma nakoupila za 30 Kč. Nakonec firma nakoupila ovoce za 20 Kč, ovoce stálo 60 Kč/Kg. Určete, za kolik korun firma průměrně zakoupila 1 Kg ovoce.²¹

Řešení:

Ještě před samotným výpočtem jsme si označili cenu ovoce za kilogram jako x_i (zjištěné hodnoty znaku X) a cenu nákupu, kterou firma za ovoce zaplatila, jako n_i (váha jednotlivých hodnot). Tyto data jsme si zobrazili v tabulce, kterou vidíme v Out[23]. Řešení příkladu v prostředí programu Mathematica vidíme na obrázku 11. Pro řešení tohoto příkladu použijeme vzorec pro vážený harmonický průměr. Tento vzorec není v programu Mathematica přednastavený, proto jsme si zavedli novou funkci VHP. Přesně formulovanou funkci VHP vidíme v In[24]. Výsledek se zobrazil ve formě zlomku, proto jsme použili výraz //N pro zobrazení numerické hodnoty výsledku. Závěr: firma průměrně nakoupila kilogram ovoce za 45,80 Kč.

```
In[23]= Grid[{{"Nákup ovoce", "xi (Kč/Kg)", "ni (Kč)"}, {1., 40, 50},
            {2., 50, 30}, {3., 60, 20}}, Frame -> All,
            Background -> {None, {LightBlue}}]

Out[23]=


| Nákup ovoce | xi (Kč/Kg) | ni (Kč) |
|-------------|------------|---------|
| 1.          | 40         | 50      |
| 2.          | 50         | 30      |
| 3.          | 60         | 20      |



In[24]= VHP[x1_, x2_, x3_, n1_, n2_, n3_] =
        (n1 + n2 + n3) / ((n1 / x1) + (n2 / x2) + (n3 / x3))
VHP[40, 50, 60, 50, 30, 20]

Out[24]=  $\frac{n1 + n2 + n3}{\frac{n1}{x1} + \frac{n2}{x2} + \frac{n3}{x3}}$ 

Out[25]=  $\frac{6000}{131}$ 

In[26]= % // N
Out[26]= 45.8015
```

Obrázek 11 Řešení příkladu č. 6

zdroj:vlastní

²¹ Srov. ŘEZANKOVÁ, H., LÖSTER, T., *Základy statistiky*, s. 22

3.1.4 Modus

Modus představuje číslo nebo hodnotu, která se v daném statistickém souboru vyskytuje nejčastěji. Modus určujeme z neroztříděného souboru, sledujeme všechny jeho hodnoty. V jednom statistickém souboru můžeme nalézt i několik modů. Pro usnadnění nalezení modu je dobré seřadit si všechny hodnoty sestupně či vzestupně. Modus značíme \hat{x} a není pro něj oficiální vzorec.²²

Výhodou této charakteristiky je, že ji můžeme použít na jakýkoliv soubor prvků, např. soubor s kvalitativními znaky. Pro větší vypovídající hodnotu je vhodné zjišťovat modus z většího množství získaných dat.

Příklad č. 7

Byl vybrán rozsah statistického souboru $N=62$ studentů druhého ročníku VŠ, v tomto souboru byl zjištěn statistický znak – výška studentů. Získané údaje vidíme na obrázku 12. Vypočítejte modus.²³

Řešení:

Řešení příkladu č. 7 vidíme na obrázku 12. Nejprve jsme si z dokumentu Microsoft Excel importovali údaje (použili jsme příkaz *Import*) a označili tyto údaje jako tabulka 1, následně jsme v In[47] vybrali ze všech importovaných údajů jen jeden sloupec s údaji o výšce studentů, se kterým budeme pracovat dále. Pro lepší orientaci máme na koncích těchto příkazů středník, příkaz se jen provede a řešení se nevypíše. Každý importovaný údaj je v samostatném seznamu, proto pro lepší práci jsme pomocí funkce *Flatten* umístili data jen do jednoho společného seznamu. Protože výška se udává nejčastěji v centimetrech, použili jsme funkci *Quantity*, která všem údajům přiřadí jednotku. Všechny zjištěné hodnoty i s jednotkami vidíme v Out[50]. Ještě předtím jsme pomocí příkazu *NumberForm* upravili vzhled dat, přidáním jednoho desetinného místa. Pro výpočet modu slouží v programu Mathematica funkce *Commonest.*, ke které jsme přidali dvě předešlé funkce, abychom dostali výsledek i s jednotkami. Závěr: studenti druhého ročníku mají nejčastější výšku 176 cm.

²² Srov. KLADIVO, P., *Základy statistiky*, s. 19

²³ Srov. ARLTOVÁ, M., aj., *Základy statistiky v příkladech*, s. 19

```

In[46]:= tabulka1 = Import["F:\\přiklady\\Příklad 7_podklady.xlsx"][[1]];
In[47]:= data = tabulka1[[Range[1, 62], Range[1, 1]]];
In[48]:= a = Flatten[data]
Out[48]= {183., 176., 169., 180., 166., 182., 191., 184., 181., 168., 190., 182., 193.,
168., 182., 185., 183., 175., 165., 176., 176., 155., 185., 187., 195., 187.,
179., 189., 165., 165., 175., 171., 160., 176., 158., 180., 176., 194.,
170., 175., 174., 174., 186., 164., 173., 172., 165., 186., 175., 185.,
188., 180., 176., 178., 175., 163., 163., 165., 182., 167., 164., 178.}

In[49]:= Quantity[{a}, "Centimeters"];
In[50]:= NumberForm[Quantity[{a}, "Centimeters"], {Infinity, 1}]
Out[50]/NumberForm=
{{ 183.0 cm , 176.0 cm , 169.0 cm , 180.0 cm , 166.0 cm , 182.0 cm , 191.0 cm ,
184.0 cm , 181.0 cm , 168.0 cm , 190.0 cm , 182.0 cm , 193.0 cm , 168.0 cm ,
182.0 cm , 185.0 cm , 183.0 cm , 175.0 cm , 165.0 cm , 176.0 cm ,
176.0 cm , 155.0 cm , 185.0 cm , 187.0 cm , 195.0 cm , 187.0 cm ,
179.0 cm , 189.0 cm , 165.0 cm , 165.0 cm , 175.0 cm , 171.0 cm ,
160.0 cm , 176.0 cm , 158.0 cm , 180.0 cm , 176.0 cm , 194.0 cm ,
170.0 cm , 175.0 cm , 174.0 cm , 174.0 cm , 186.0 cm , 164.0 cm ,
173.0 cm , 172.0 cm , 165.0 cm , 186.0 cm , 175.0 cm , 185.0 cm ,
188.0 cm , 180.0 cm , 176.0 cm , 178.0 cm , 175.0 cm , 163.0 cm ,
163.0 cm , 165.0 cm , 182.0 cm , 167.0 cm , 164.0 cm , 178.0 cm }}

In[51]:= NumberForm[Quantity[Commonest[a], "Centimeters"], {Infinity, 1}]
Out[51]/NumberForm=
{ 176.0 cm }

```

Obrázek 12 Řešení příkladu č. 7

zdroj: vlastní

3.1.5 Medián

„Medián je prvek řady (hodnot sledovaného znaku), uspořádané v neklesajícím (rostoucím) pořadí, který ji rozděluje na dvě části v tom smyslu, že polovina prvků této řady má menší hodnotu znaku a polovina má větší hodnotu znaku, než je hodnota mediánu. Jinými slovy lze prohlásit, že za medián považujeme hodnotu, která nám dělí vzestupně seřazené hodnoty statistického souboru na dvě stejné poloviny. Označujeme ho \tilde{x} nebo Me.“²⁴

Medián má několik výhod i nevýhod. Hlavní výhodou je, že na rozdíl od aritmetického průměru, extrémní hodnoty (mim. a max.) nemají na medián vliv. Medián můžeme použít pro pořadový statistický znak. Další výhodou je, že medián zachycuje polohu hodnot lépe než průměr. Nevýhodou je, že z mediánu ne učiníme závěr o součtu

²⁴ KLADIVO, P., *Základy statistiky*, s. 20

hodnot a medián neumožňuje výpočet celkové charakteristiky souboru na základě dílčích charakteristik.²⁵

Pro určení mediánu jsou k dispozici dvě rovnice. Máme daný rozsah souboru N , který má hodnoty x_1, x_2, \dots, x_N . Pokud je počet hodnot lichý, pak je medián hodnota, která se nachází na pozici:

$$\tilde{x} = x_{\left(\frac{N+1}{2}\right)}, \quad (7)$$

pokud je počet hodnot sudý, medián počítáme jako aritmetický průměr prostředních hodnot:

$$\tilde{x} = \frac{x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}}{2}. \quad (8)$$

Také můžeme říci, že medián závisí na pořadí hodnot znaku v souboru, naopak hodnota mediánu nezávisí na všech hodnotách znaku.²⁶

Příklad č. 8

Český statistický úřad zveřejnil následující data o počtu cizinců, kteří měli na území České republiky trvalý pobyt v letech 2006-2014. Charakterizujte tento statistický soubor vhodně zvolenou charakteristikou polohy. Data vidíme v tabulce 4.

Tabulka 4 Počet cizinců s trvalým pobytem

Rok	Počet cizinců
2006	56 797
2007	64 352
2008	66 754
2009	66 855
2010	69 816
2011	75 249
2012	80 844
2013	99 467
2014	249 856

zdroj: ČSU, www.czso.cz/csu/czso/ceska-republika-od-roku-1989-v-cislech

²⁵ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 108

²⁶ Srov. tamtéž

Řešení:

Jak vidíme v tabulce 4, skoro všechny hodnoty jsou v řádu desetitisíců, jen poslední hodnota je extrémní, je v řádu statisíců. Pokud máme charakterizovat tyto data, potřebuje charakteristiku, která není ovlivněna extrémními hodnotami. Tou charakteristikou je medián. Ještě před začátkem samotného výpočtu mediánu, vidíme na obrázku 13, jaký výsledek dostaneme, když použijeme aritmetický průměr. Aritmetický průměr je ovlivněn extrémními hodnotami, proto je výsledek v Out[14] zkreslující (zadané hodnoty jsou menší než výsledek). Naopak pokud zjišťujeme medián, vyjde nám v Out[15] hodnota 69 816, což je přijatelný výsledek, vzhledem k zadaným hodnotám. Protože je počet hodnot lichý, výsledek je přesně prostřední hodnota. V případě sudého počtu hodnot nám vyjde zcela nová hodnota, kterou mezi zadanými údaji nenajdeme, tak jak to vidíme v Out[16]. Zde jsme počítali medián z 8 vybraných hodnot. Závěr: pokud se vyskytuje v souboru dat jedna extrémní hodnota, je lepší použít k charakteristice souboru medián. V tomto případě nám medián vyšel 69 816, což lze označit na prostřední hodnotu zadaného souboru dat.

```
In[13]:= Mean[{56 797, 64 352, 66 754, 66 855, 69 816, 75 249, 80 844, 99 467, 249 856}]
Out[13]=  $\frac{829\,990}{9}$ 

In[14]:= % // N
Out[14]= 92 221.1

In[15]:= Median[{56 797, 64 352, 66 754, 66 855, 69 816, 75 249, 80 844, 99 467, 249 856}]
Out[15]= 69 816

In[16]:= Median[{64 352, 66 754, 66 855, 69 816, 75 249, 80 844, 99 467, 249 856}]
Out[16]=  $\frac{145\,065}{2}$ 

In[17]:= % // N
Out[17]= 72 532.5
```

Obrázek 13 Řešení příkladu č. 8

zdroj vlastní

3.2 Charakteristiky variability

V předchozí podkapitole jsme se zabývali charakteristikami polohy. Ty nám pouze poskytovali informace o poloze statistického souboru na číselné ose, zjišťovali jsme převážně střední hodnoty. Někdy tyto informace nestačí a my potřebujeme vědět, jak moc jsou hodnoty v souboru rozptýleny kolem střední hodnoty. K tomu nám slouží charakteristiky variability. Tyto charakteristiky ukazují rozmístění hodnot proměnné okolo střední hodnoty, jaké jsou vzdálenosti jednotlivých hodnot od průměru. Charakteristiky variability také zjišťují, zda jsou v souboru extrémní hodnoty.

Mezi charakteristiky variability řadíme: variační rozpětí, rozptyl, směrodatná odchylka, variační koeficient, průměrná odchylka. Všechny tyto pojmy si vysvětlíme v následujících kapitolách, kde také najdete několik příkladů s touto tematikou.

Výše uvedené charakteristiky můžeme rozdělit do dvou kategorií. První kategorie se nazývá charakteristiky absolutní variability. Tyto charakteristiky jsou vyjádřeny ve stejných měrných jednotkách jako hodnoty zkoumaného znaku. Do této kategorie řadíme: variační rozpětí, průměrná odchylka, rozptyl a směrodatná odchylka. Druhá kategorie se nazývá charakteristiky relativní variability. Tyto charakteristiky jsou bezrozměrné nebo se udávají v procentech. Používáme je v případě, když chceme porovnávat variabilitu různých statistických souborů vyjádřených v různých měrných jednotkách. V této kategorii se nachází variační koeficient.²⁷

3.2.1 Variační rozpětí

První, asi nejjednodušší ukazatel variability, je variační rozpětí. Variační rozpětí značíme písmenem R a vypočítáme ho jako rozdíl maximální a minimální hodnoty ve sledovaném souboru:

$$R = x_{\max} - x_{\min} . \quad (9)$$

Tento ukazatel je závislý pouze na dvou extrémních hodnotách, proto nemusí mít dostatečně vypovídající hodnotu. Ukazatel slouží především k poskytnutí první základní informací o statistickém souboru.²⁸

Variační rozpětí má své výhody i nevýhody. Mezi výhody řadíme snadnost a rychlost výpočtu a jednoduchý výklad (můžeme říci, že variační rozpětí nám stanoví délku intervalu, v kterém se pohybují hodnoty sledovaného znaku). Naopak nevýhodou je, že variační rozpětí pracuje s extrémními hodnotami, pokud se v souboru vyskytnou

²⁷ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 117

²⁸ Srov. KLADIVO, P., *Základy statistiky*, s. 22

jen jedna extrémní hodnota, potom má variační rozpětí značnou velikost. Variační rozpětí nám necharakterizuje variabilitu hodnot uvnitř variačního rozpětí.²⁹

3.2.2 Průměrná odchylka

Průměrná odchylka zkoumá míru odlišnosti jednotlivých hodnot znaků od střední hodnoty (průměr, modus, medián). Průměrnou odchylku značíme \bar{d} (k tomuto symbolu přidáváme dolní index, podle toho od které střední hodnoty se odchylka počítá). Obecný vzorec pro výpočet průměrné odchylky vidíme níže:

$$\bar{d}_a = \frac{\sum_{i=1}^n |x_i - a|}{n}, \quad (10)$$

kde x_i je konkrétní zjištěná hodnota statistického znaku, a je zvolená charakteristika polohy, n je počet hodnot statistického znaku. Můžeme říci, že průměrná odchylka se spočítá jako aritmetický průměr rozdílů všech hodnot znaku od střední hodnoty.³⁰

3.2.3 Rozptyl a směrodatná odchylka

Tyto dvě charakteristiky spolu úzce souvisejí, budeme se jim proto věnovat ve společné podkapitole. Velikost těchto charakteristik je závislá na všech hodnotách statistického znaku. Obě tyto charakteristiky měří rozptýlenost dat kolem aritmetického průměru a proměnlivost odchylek jednotlivých hodnot znaku.³¹

Rozptyl značíme s^2 a je definován jako aritmetický průměr druhých mocnin odchylek hodnot znaku od aritmetického průměru. Vzorec pro rozptyl vidíme níže:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \quad (11)$$

kde x_i je konkrétní zjištěná hodnota statistického znaku, \bar{x} je aritmetický průměr celého souboru, n je počet hodnot statistického znaku. Rozptyl se používá jako míra teoretická, je nejdůležitější charakteristikou, ale méně používanou.³²

²⁹ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 113

³⁰ Srov. KLADIVO, P., *Základy statistiky*, s. 22-23

³¹ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 114

³² Srov. KLADIVO, P., *Základy statistiky*, s. 23

Směrodatná odchylka se značí s a je definována jako odmocnina z rozptylu:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}. \quad (12)$$

Tuto charakteristiku používáme v praxi častěji, je to nejvhodnější charakteristika. Její výhodou je, že se uvádí ve stejných měrných jednotkách jako zkoumaný statistický znak, ale i ona je značně ovlivněna extrémními hodnotami.³³

Příklad č. 9

Pro účely evidence cen pohonných hmot v Praze, byl vytvořen statistický soubor se dvěma argumenty, u vybraných čerpacích stanic jsme zjišťovali ceny benzínu a nafty (tabulka 5 na obrázku 14). Charakterizujte variabilitu ceny těchto pohonných hmot.

```

In[26]:= tab1 = Import["F:\\přiklady\\Příklad 9 podklady.xlsx"][[1]];

In[27]:= tab1 // TableForm
Out[27]/TableForm=

```

Název čerpací stanice	Poloha	Cena benzínu (Kč/l)	Cena nafty (Kč/l)
EuroOil	Praha 3	28.8	26.6
OMV	Praha 8	29.5	27.9
Benzina Plus	Praha 7	29.2	27.5
Pap Oil	Praha 5	28.4	26.9
Star (Benzina)	Praha 9	28.3	26.7
Shell	Praha 3	30.5	26.9
Totoil	Praha 3	27.9	25.4
Medos	Praha 3	28.5	26.5
MOL	Praha 10	29.4	27.8
Bruto a.s.	Praha 10	28.9	26.9

```

In[36]:= tab2 = Grid[tab1, Dividers -> {3 -> True, 2 -> True}, Frame -> True,
  Spacings -> {Automatic, 1}];

In[38]:= Labeled[tab2, "Tabulka 5 Ceny pohonných hmot", Top,
  LabelStyle -> Directive[Bold, FontFamily -> "Helvetica"]]
Out[38]=

```

Tabulka 5 Ceny pohonných hmot			
Název čerpací stanice	Poloha	Cena benzínu (Kč/l)	Cena nafty (Kč/l)
EuroOil	Praha 3	28.8	26.6
OMV	Praha 8	29.5	27.9
Benzina Plus	Praha 7	29.2	27.5
Pap Oil	Praha 5	28.4	26.9
Star (Benzina)	Praha 9	28.3	26.7
Shell	Praha 3	30.5	26.9
Totoil	Praha 3	27.9	25.4
Medos	Praha 3	28.5	26.5
MOL	Praha 10	29.4	27.8
Bruto a.s.	Praha 10	28.9	26.9

Obrázek 14 Data k příkladu č. 9

zdroj dat: http://www.mbenzin.cz/Ceny-benzinu-a-nafty-Praha-M_

zdroj obrázku: vlastní

³³ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 116

Řešení:

Řešení příkladu vidíme na obrázku 15 (viz následující strana). Nejprve jsme si pomocí příkazu *Range* vybrali z tab1 požadované hodnoty (ceny benzínu x_i a nafty y_i). Máme za úkol charakterizovat variabilitu těchto dat, proto budeme u souboru počítat charakteristiky variability. Prvotní informace o souboru dat jsme získali z výpočtu variačního rozpětí. V programu Mathematica není tento vzorec přednastavený, proto jsme si zavedli novou funkci *R*, která je definovaná jako maximální hodnota v souboru mínus minimální hodnota v souboru. Velikost intervalu, na kterém se pohybují ceny benzínu, je 2,6 Kč, u nafty je to 2,5 Kč. Dále jsme počítali průměrnou odchylku od střední hodnoty (zde od průměru). Využili jsme přednastavenou funkci *MeanDeviation*. Zjistili jsme, že průměrné odchýlení hodnot v souboru od aritmetického průměru je v případě benzínu 0,568 Kč a v případě nafty 0,494 Kč. Další důležitou charakteristikou je rozptyl, výpočet přes funkci *Variance*. I zde je větší vypočítaná hodnota u benzínu (přibližně 0,56) než u nafty (přibližně 0,52), ceny benzínu jsou tedy více rozptýleny na číselné ose. Poslední charakteristikou, kterou jsme zjišťovali, je směrodatná odchylka. Zde jsme použili funkci *Sqrt*, v kombinaci s funkcí *Variance* zjistíme odmocninu z rozptylu. U benzínu to je 0,75 Kč, u nafty 0,72 Kč.

Závěr: ze zjištěných hodnot je zřejmé, že větší variabilita (zjišťovaná rozptylem nebo směrodatnou odchylkou) je u ceny benzínu, všechny výsledky charakteristik variability jsou vyšší než výsledky u nafty. Průměrná cena benzínu je 28,94 Kč/l se směrodatnou odchylkou 0,75 Kč, průměrná cena nafty je 26,91 Kč/l se směrodatnou odchylkou 0,72 Kč.

```

In[5]:= xi = tab1[[Range[2, 11], Range[3, 3]]]
Out[5]= {{28.8}, {29.5}, {29.2}, {28.4}, {28.3}, {30.5}, {27.9}, {28.5}, {29.4}, {28.9}}

In[6]:= yi = tab1[[Range[2, 11], Range[4, 4]]]
Out[6]= {{26.6}, {27.9}, {27.5}, {26.9}, {26.7}, {26.9}, {25.4}, {26.5}, {27.8}, {26.9}}

In[7]:= Rx = Quantity[Max[xi] - Min[xi], "CzechKoruny"]
Out[7]= Kč2.6

In[8]:= Ry = Quantity[Max[yi] - Min[yi], "CzechKoruny"]
Out[8]= Kč2.5

In[9]:=  $\overline{d_{xi}}$  = MeanDeviation[xi]
Out[9]= {0.568}

In[10]:=  $\overline{d_{yi}}$  = MeanDeviation[yi]
Out[10]= {0.494}

In[11]:= Round[Variance[xi], 0.01]
Out[11]= {0.56}

In[12]:= Round[Variance[yi], 0.01]
Out[12]= {0.52}

In[13]:= Round[Sqrt[Variance[xi]], 0.01]
Out[13]= {0.75}

In[14]:= Round[Sqrt[Variance[yi]], 0.01]
Out[14]= {0.72}

In[17]:= Mean[xi];
In[18]:= Mean[yi];

```

Obrázek 15 Řešení příkladu č. 9

zdroj: vlastní

3.2.4 Variační koeficient

Variační koeficient značíme V a je formulován jako poměr směrodatné odchylky a aritmetického průměru:

$$V = \frac{S}{\bar{x}}. \quad (13)$$

Tento ukazatel je bezrozměrný, výsledkem je desetinné číslo, po vynásobení vzorce stem dostaneme variabilitu v procentech. Variační koeficient je nejpoužívanější relativní mírou variability, používá se při porovnávání variability dvou různých souborů.³⁴

³⁴ Srov. KLADIVO, P., *Základy statistiky*, s. 24

Příklad č. 10

Určete, ve které firmě byla výroba rovnoměrnější, pokud známe jejich produkci během deseti dní. Firma X počítá své výrobky po kusech, firma Y vykazuje svou produkci v tunách. Produkci obou firem každý den vidíme v tabulce 6.³⁵

Tabulka 6 Zadání k příkladu č. 10

Den	Firma X (ks)	Firma Y (tuny)
1	1000	6
2	2000	6
3	2000	5
4	3000	8
5	2000	9
6	4000	4
7	2000	4
8	1000	6
9	2000	5
10	4000	7

zdroj: ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 117

Řešení:

Protože je produkce firem vyjádřena v různých měrných jednotkách, musíme k porovnání rovnoměrnosti výroby použít variační koeficient. Budeme zjišťovat variační koeficienty pro každou firmu zvlášť a potom je porovnáme. Výpočet vidíme na obrázku 16, kde jsme si definovali složené příkazy Vx a Vy. Hlavní jádro příkazu je výpočet variačního koeficientu jako směrodatná odchylka dělená aritmetickým průměrem. Abychom dostali variabilitu v procentech, vynásobili jsme tento vzorec stem, ostatní příkazy jsou jen úprava výsledku do přijatelné podoby. Závěr: rovnoměrnější výroba je ve firmě Y, protože její variační koeficient vyšel méně (27,22%) než variační koeficient firmy X (49,06%).

```
In[28]:= xi = {1000, 2000, 2000, 3000, 2000, 4000, 2000, 1000, 2000, 4000};
In[29]:= yi = {6, 6, 5, 8, 9, 4, 4, 6, 5, 7};
In[30]:= Vx = Quantity[N[(Sqrt[Variance[xi]] / Mean[xi]) * 100, 4], "percent"]
Out[30]= 46.06%
In[31]:= Vy = Quantity[N[(Sqrt[Variance[yi]] / Mean[yi]) * 100, 4], "percent"]
Out[31]= 27.22%
```

Obrázek 16 Řešení příkladu č. 10

vzor: vlastní

³⁵ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 117

3.3 Charakteristiky šikmosti

Charakteristiky šikmosti se zabývají otázkou, jestli jsou hodnoty proměnné rozloženy kolem zvoleného středu (např. aritmetický průměr) souměrně nebo asymetricky (např. zešikmeny na jednu stranu). Charakteristiky šikmosti se používají méně často (převážně ve spojení s charakteristikami špičatosti) a poskytují doplňující informace o statistickém souboru. Pomocí této charakteristiky jsme schopni odhadnout tvar rozdělení četností. Mezi charakteristiky šikmosti řadíme charakteristiku zvanou koeficient šikmosti, kterou se budeme zabývat na následujících řádcích.³⁶

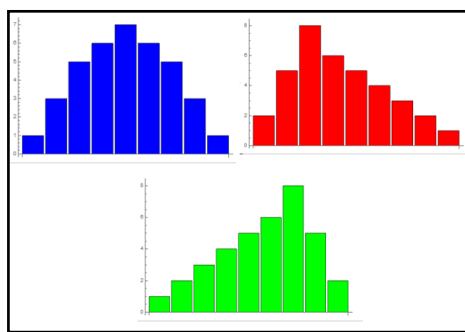
3.3.1 Koeficient šikmosti

Koeficient šikmosti značíme α_x a vypočítáme ho podle vzorce:

$$\alpha_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n s^3}, \quad (14)$$

kde x_i je konkrétní zjištěná hodnota statistického znaku, \bar{x} je aritmetický průměr celého souboru, n je počet hodnot statistického znaku a s je směrodatná odchylka. Podle toho, jak velký koeficient šikmosti vychází, definujeme tři základní rozložení četností hodnot statistického znaku:

- symetrické, pokud $\alpha_x = 0$, hodnoty jsou rovnoměrně rozloženy kolem střední hodnoty, grafické znázornění vidíme na obrázku 17 - modrý graf
- kladně zešikmené, pokud $\alpha_x > 0$, v souboru se nacházejí více malé hodnoty a graficky se to projeví pozvolným klesáním (viz obrázek 17 - červený graf),
- záporně zešikmené, pokud $\alpha_x < 0$, v souboru jsou koncentrovány vysoké hodnoty, graficky se tato situace projeví pozvolnějším růstem hodnot (viz obrázek 17 - zelený graf).³⁷



Obrázek 17 Grafické znázornění rozložení četností

zdroj: vlastní, vytvořeno v programu Mathematica

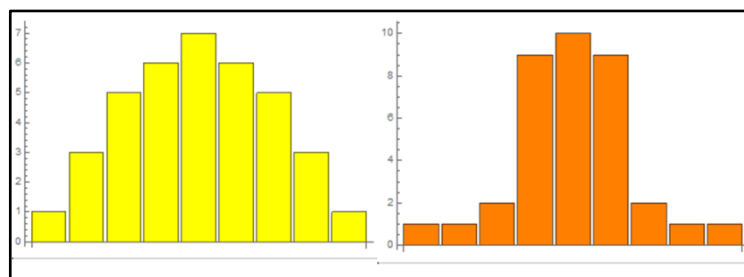
³⁶ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 119

³⁷ Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 31.

3.4 Charakteristiky špičatosti

„Charakteristiky špičatosti měří stupeň koncentrace hodnot kolem středu ve srovnání s četností ostatních hodnot.“³⁸

Rozlišujeme dvě situace: pokud je v souboru stejná koncentrace prostředních hodnot a ostatních hodnot, potom se špičatost projevuje plochým tvarem rozdělení četností. V takovém to souboru s nízkou špičatostí můžeme očekávat hodnoty velmi vzdálené od středu. Druhá situace je přesně opačná, v souboru se vyskytuje větší koncentrace prostředních hodnot než ostatních hodnot, rozdělení má špičatý tvar. Tyto dvě situace můžeme graficky znázornit podle obrázku 18. Žlutý graf představuje plochý tvar rozdělení četností, oranžový graf představuje špičatý tvar.³⁹



Obrázek 18 Grafická interpretace špičatosti

zdroj: vlastní, vytvořeno v programu Mathematica

3.4.1 Koeficient špičatosti

Koeficient špičatosti značíme β_x a vypočítáme ho podle vzorce:

$$\beta_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{ns^4} - 3, \quad (15)$$

kde x_i je konkrétní zjištěná hodnota statistického znaku, \bar{x} je aritmetický průměr celého souboru, n je počet hodnot statistického znaku a s je směrodatná odchylka. Čím je koeficient špičatosti vyšší, tím je rozdělení četností strmější, je větší špičatost rozdělení a v souboru je více hodnot blízkých střední hodnotě. Koeficient špičatosti je bezrozměrné číslo.⁴⁰

³⁸ ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 122

³⁹ Srov. ŠALOUNOVÁ, D., *Úvod do pravděpodobnosti a statistiky*, s. 122

⁴⁰ Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 32.

Příklad č. 11

Charakterizujte statistický soubor, který jsme získali od Českého statistického úřadu o počtu dětí v mateřských školách v České republice od školního roku 2005/06 až do 2014/15. Použijte charakteristiky špičatosti a šikmosti, dokažte výsledky i graficky. Data vidíme v tabulce 7.

Tabulka 7 Počty dětí v mateřských školách

Školní rok	Počet dětí	Školní rok	Počet dětí
2005/06	282 183	2010/11	328 612
2006/07	285 419	2011/12	342 521
2007/08	291 194	2012/13	354 340
2008/09	301 620	2013/14	363 568
2009/10	314 008	2014/15	367 603

zdroj: ČSU, <https://www.czso.cz/csu/czso/ceska-republika-od-roku-1989-v-cislech#12>

Řešení:

Řešení příkladu v programu Mathematica vidíme na obrázku 19. Pro výpočet koeficientu šikmosti jsme použili předdefinovanou funkci *Skewness*, výsledek vyšel kladné číslo 0,090, což značí, že v souboru dat se nacházejí spíše malé hodnoty, hodnoty nejsou symetrické, můžeme očekávat kladné zešikmení. Dále jsme počítali koeficient špičatosti pomocí příkazu *Kurtosis*. Výsledkem je záporné číslo -1,523, předpokládáme, že soubor hodnot má plochý tvar, hodnoty nejsou koncentrovány kolem středu. Pro ověření předpokladů jsme sestrojili v programu Mathematica sloupcový graf, který vidíme na obrázku 20. Z obrázku je zřejmé, že soubor hodnot má opravdu plochý tvar, hodnoty nejsou symetrické a převládají nízké hodnoty.

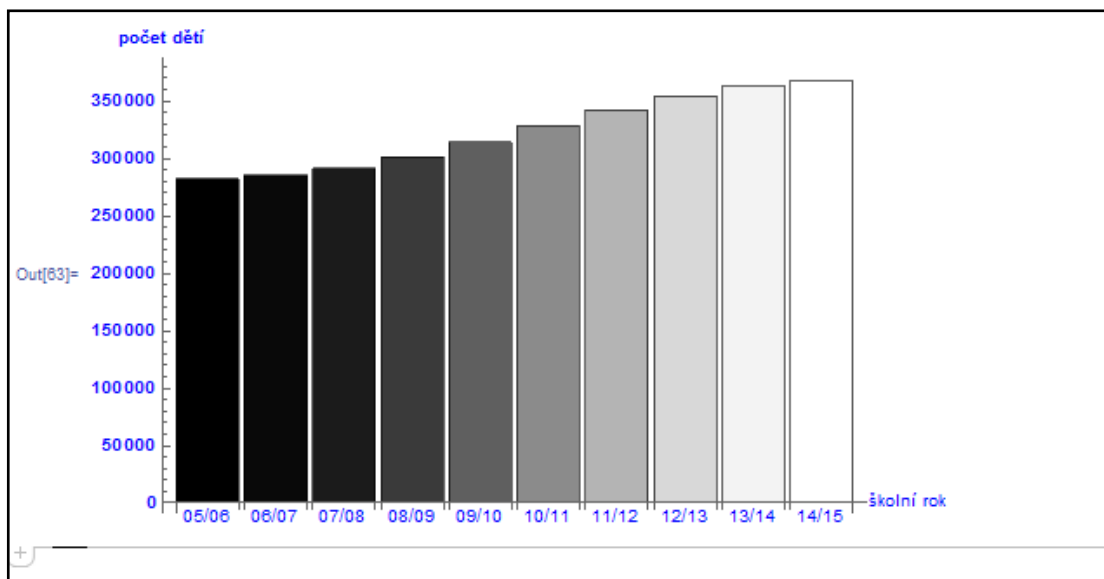
```
In[18]:= deti = {282183, 285419, 291194, 301620, 314008, 328612, 342521,
              354340, 363568, 367603};

In[8]:= Skewness[deti] // N
Out[8]= 0.0901207

In[17]:= Kurtosis[deti] - 3 // N
Out[17]= -1.52323
```

Obrázek 19 Řešení příkladu č. 11

zdroj: vlastní



Obrázek 20 Graf k příkladu č. 11 vyjadřující počet dětí v mateřských školách v jednotlivých školních rocích

zdroj: vlastní

4 ROZDĚLENÍ ČETNOSTÍ A JEJICH GRAFICKÉ VYJÁDŘENÍ

Poslední kapitola této práce má název rozdělení četností a jejich grafické vyjádření. První část této kapitoly bude věnovaná rozdělení četností, ve druhé části se budeme zabývat grafickým vyjádřením četností, představíme si různé typy grafů a pro lepší názornost si je v programu Mathematica vykreslíme.

„Četností rozumíme počet prvků se stejnou hodnotou statistického znaku (každý statistický soubor tak generuje své rozdělení četností) nebo četností myslíme počet prvků s hodnotami znaku patřícími do určitého intervalu (nebo třídy)-pak se bavíme o tzv. skupinovém (intervalovém) rozdělení četností.“⁴¹

Četnosti rozdělujeme do tří základních kategorií. Můžeme se setkat s absolutní četností, relativní četností a kumulativní četností. Absolutní četnost značíme n_i a je to číslo, které představuje počet opakování konkrétní hodnoty znaku v statistickém souboru. Relativní četnost zjistíme jako podíl jednotlivé absolutní četnosti n_i a rozsahu souboru n . Po vynásobení tohoto zlomku stem, získáme informaci, kolik procent z celkového souboru tvoří příslušná hodnota znaku. Kumulativní četnost může být absolutní i relativní a tato četnost vzniká postupným načítáním četností (relativních nebo absolutních). Kumulativní absolutní četnost značíme N_i , kumulativní relativní četnost P_i .⁴²

Ve statistice hrají svou roli pro zpracovávání údajů nejen vzorce pro výpočet a tabulky, ale i grafy. Statistické grafy jsou jednou z možností, jak rychle a přehledně prezentovat data. Při práci se statistickými výsledky máme k dispozici velké množství dat, tyto data musíme roztrždit, vyhodnotit a představit výsledky. K tomu všemu nám slouží ve statistice různé typy grafů, které si na následujících řádcích popíšeme a uvedeme příklad vytvořený v programu Mathematica.

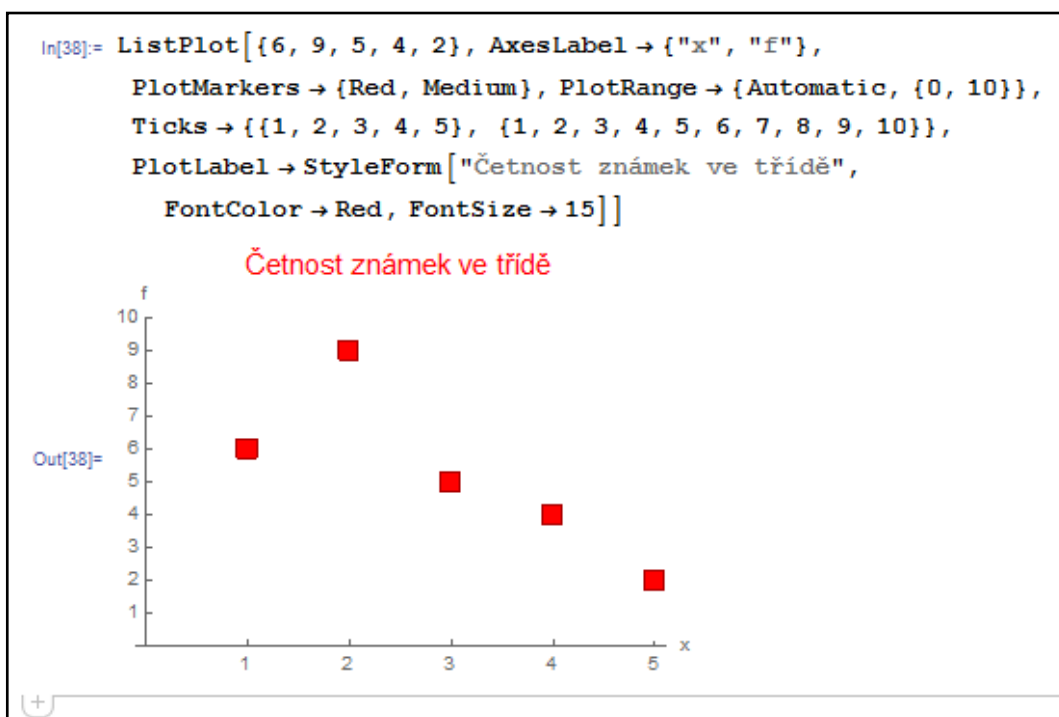
⁴¹ KLADIVO, P., *Základy statistiky*, s. 11

⁴² Srov. LÖSTER, T., ŘEZANKOVÁ, H., LANGHAMROVÁ, J., *Statistické metody a demografie*, s. 20.

4.1 Bodový graf rozložení četnosti

Bodový graf je nejjednodušší vyjádření četností, jak relativní i absolutní. Bodový graf vznikne v pravoúhlém souřadnicovém systému, do tohoto systému vynášíme body se souřadnicemi [hodnota znaku, četnost], na osu x nanášíme hodnoty znaku, na osu y četnosti. Čím je bod v grafu položen výše, tím je četnost hodnoty vyšší.

Příklad bodového grafu vidíme na obrázku 21. Tento graf byl vytvořen v programu Mathematica a představuje rozložení známek z testu mezi studenty. Vidíme, že nejvíce studentů získalo z testu známku 2. Pro tento graf jsme použili příkaz *ListPlot*, který vykreslí samostatné body. Dále, aby byl graf esteticky přijatelný, jsme použili několik příkazů doplňující vzhled grafu. Byli to tyto příkazy: *AxesLabel* přiřadí popisky k osám, *PlotMarkers* mění vzhled samotných bodů, příkaz *Ticks* umístí na osy vybrané hodnoty, *PlotRange* upravuje rozsah os u grafu, *PlotLabel* přiřadí popisek celému grafu.



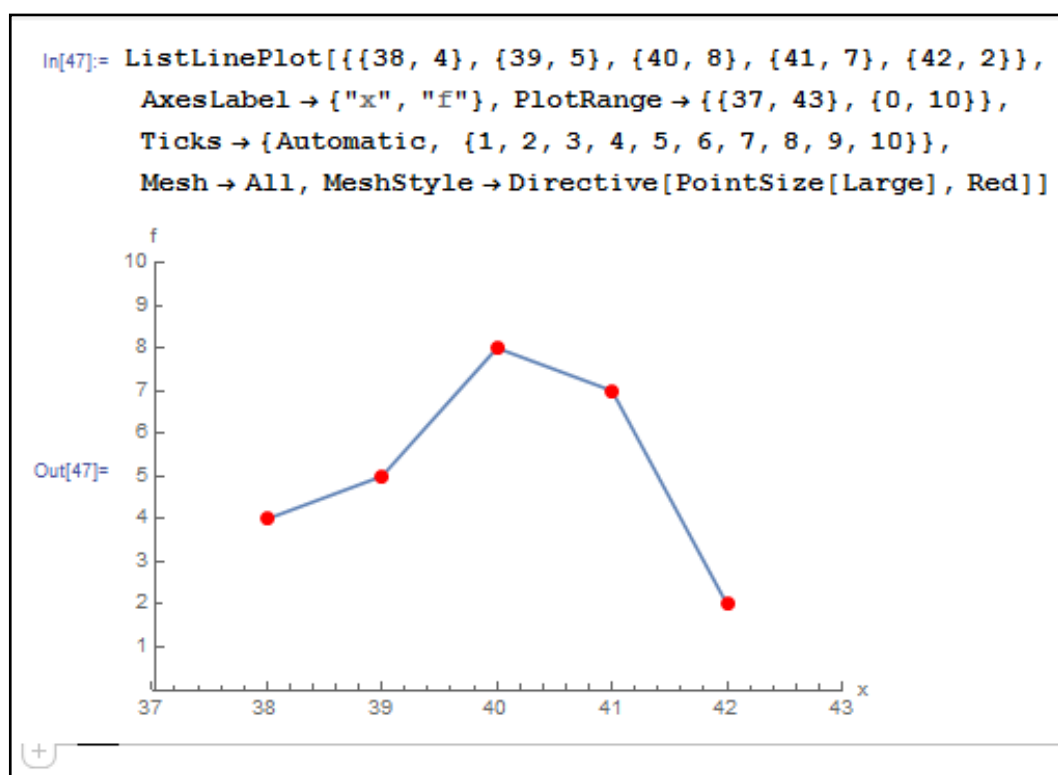
Obrázek 21 Četnost známek z testu

zdroj:vlastní

4.2 Polygon rozložení četností

Polygon je spojnicový typ grafu, tento graf sestojíme stejně jako bodový graf a následně tyto body spojíme úsečkami. Vznikne tak lomená čára, která představuje hledaný polygon. Příklad polygonu vidíme na obrázku 22. Tento graf vyjadřuje, kolik párů obuvi v různých velikostech se prodalo v obchodě s obuví během 1 měsíce.

K vytvoření tohoto grafu jsme použili příkaz *ListLinePlot*, který vykreslí lomenou čáru. Některé příkazy jsou stejné jako u předchozího grafu, my si vysvětlíme jen ty nové. K zobrazení červených bodů v grafu jsme použili příkaz *Mesh*, u kterého můžeme nastavit pomocí příkazu *MeshStyle* další vlastnosti (barva bodů, velikost).



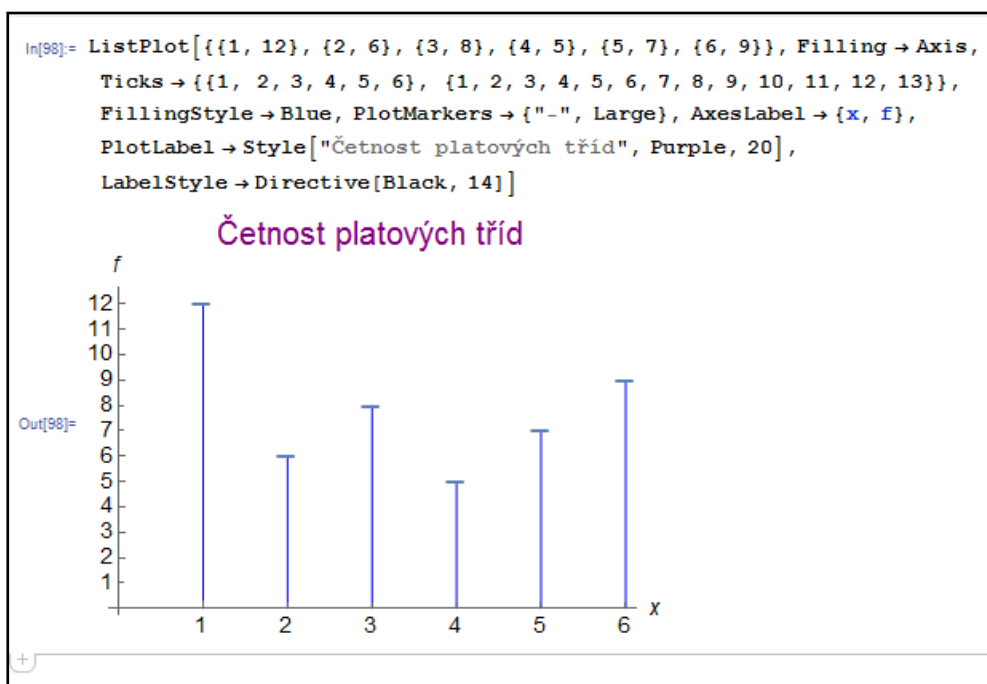
Obrázek 22 Četnost prodeje obuvi

zdroj: vlastní

4.3 Úsečkový graf

Úsečkový graf může znázorňovat rozložení absolutních a relativních četností, sestrojíme ho opět v pravoúhlém souřadnicovém systému. Na ose x se nacházejí hodnoty znaku x, na ose y odpovídající četnost. Z bodů na ose x sestrojujeme úsečky rovnoběžné s osou y, tyto úsečky mají stejnou délku jako příslušné četnosti. Tak vznikne úsečkový graf. Úsečkový graf můžeme vidět na obrázku 23. Tento graf vyjadřuje počty zaměstnanců ve firmě, kteří mají stejnou platovou třídu.

V programu Mathematica jsme použili příkaz `ListPlot`, doplněný příkazem `Filling`, který vykreslil jednotlivé úsečky. Příkaz `FillingStyle` udává vlastnosti jednotlivým úsečkám jako je barva, tloušťka nebo typ čáry. V celém grafu jsme změnili pomocí příkazu `LabelStyle` velikost a barvu písma. Pomocí `PlotLabel` jsme přidali popisek, který jsme také upravili pomocí příkazu `Style`.



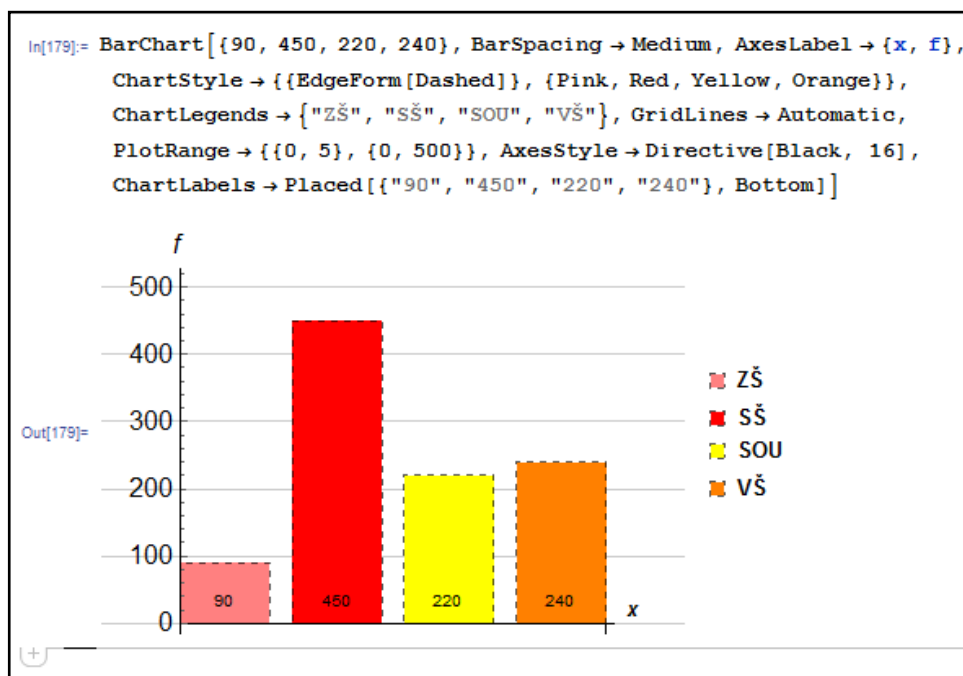
Obrázek 23 Rozložení platových tříd mezi zaměstnanci

zdroj:vlastní

4.4 Sloupcový graf rozložení četností

Sloupcový graf má podobný vzhled jako úsečkový graf, jen místo úseček jsou v grafu zobrazeny sloupce. Výška těchto sloupců vyjadřuje četnosti jednotlivých hodnot. Opět na osu x zaznamenáváme hodnoty znaku, na ose y jsou četnosti. Poloha sloupců v grafu může být rozdílná, buď jsou umístěny ve svislé poloze, nebo ve vodorovné, sloupce grafu se nikdy nedotýkají. Často se sloupcový graf používá k vyjádření četností různých kategorií, kdy jeden sloupec zastupuje vždy jednu kategorii. Opět platí čím je vyšší sloupec, tím je hodnota zastoupena v souboru více. Sloupcový graf můžeme použít jak při vyjádření absolutních i relativních četností.

Ukázka sloupcového grafu je na obrázku 24. Graf vyjadřuje nejvyšše dosažené vzdělání mezi 1000 dotázanými občany. Graf je vytvořen v programu Mathematica a pro vytvoření grafu jsme použili příkaz *BarChart*. Dále jsme použili příkazy, které řadíme do kategorie možností grafu. *BarSpacing* určuje šířku mezi jednotlivými sloupci, *ChartStyle* upravuje vzhled jednotlivých sloupců (barvy, styl, rámování). *ChartLegends* přiřadí ke grafu legendu. Příkaz *GridLines* vloží do grafu vodící linky, *ChartLabels* umístí na jednotlivé sloupce popisky. *AxesStyle* upravuje vzhled os (barva, velikost písma, styl písma).



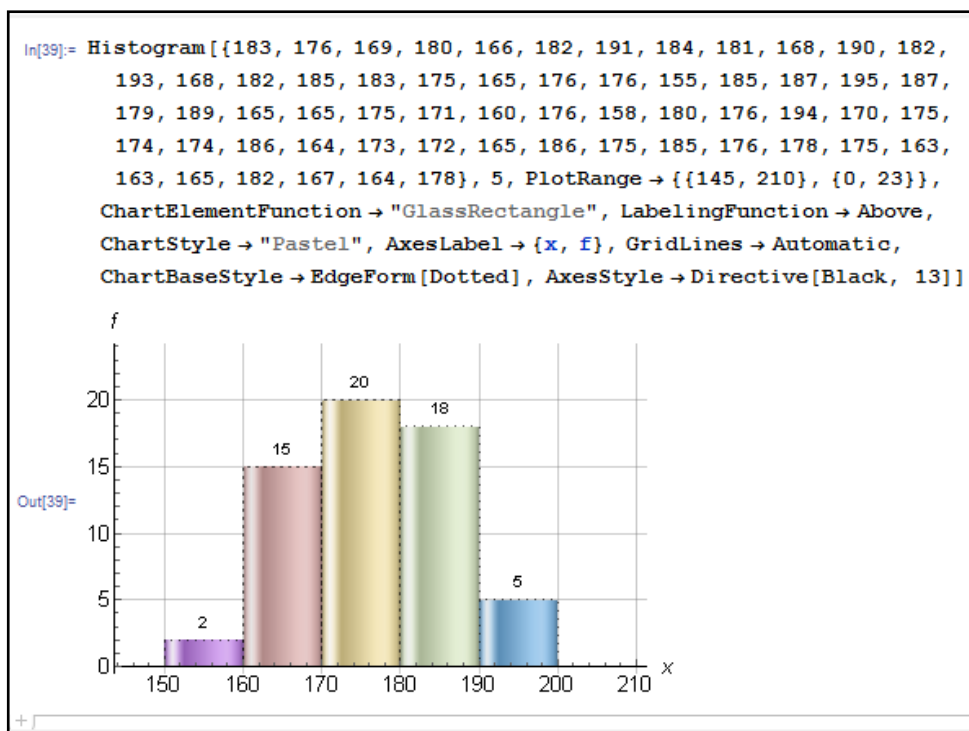
Obrázek 24 Četnost dosaženého vzdělání

zdroj:vlastní

4.5 Histogram rozdělení četností

Histogram je speciální obměna sloupcového grafu. Histogram použijeme, když máme hodnoty znaku roztržené do intervalů. Na ose x jsou umístěny jednotlivé intervaly, na osu y opět vynášíme jednotlivé četnosti (absolutní i relativní). Nad každým intervalem sestrojíme sloupec, který svou výškou odpovídá velikosti četnosti. Sloupce v histogramu jsou vždy ve vertikálním směru, šířka sloupců se mění podle šířky intervalu. Mezi jednotlivými sloupci není nikdy mezera.

Vzor histogramu vidíme na obrázku 25. Tento graf vyjadřuje rozložení dat z příkladu č. 7 (data o výšce studentů). U studentů jsme nejprve zjišťovali jejich výšku, získaná data jsme rozdělili do intervalů. Výsledek šetření vidíme v histogramu, nejčastější naměřená výška u studentů byla v intervalu 170-180 cm. Pro vytvoření histogramu v programu Mathematica jsme použili příkaz *Histogram* a zvolili jsme si 5 intervalů, do kterých se data mají rozdělit. Další příkazy jsou podobné jako u předchozích grafů, nově použité jsou zde příkazy *ChartElementFunction*, který upravuje styl sloupců. Tento příkaz kombinujeme s *ChartStyle*, který mění barvu sloupců. Nový je zde i příkaz *LabelingFunction*, který zobrazí jednotlivé četnosti přímo u sloupců. Zvolili jsme i jiné orámování sloupců, použili jsme vlastnost *Dotted*.



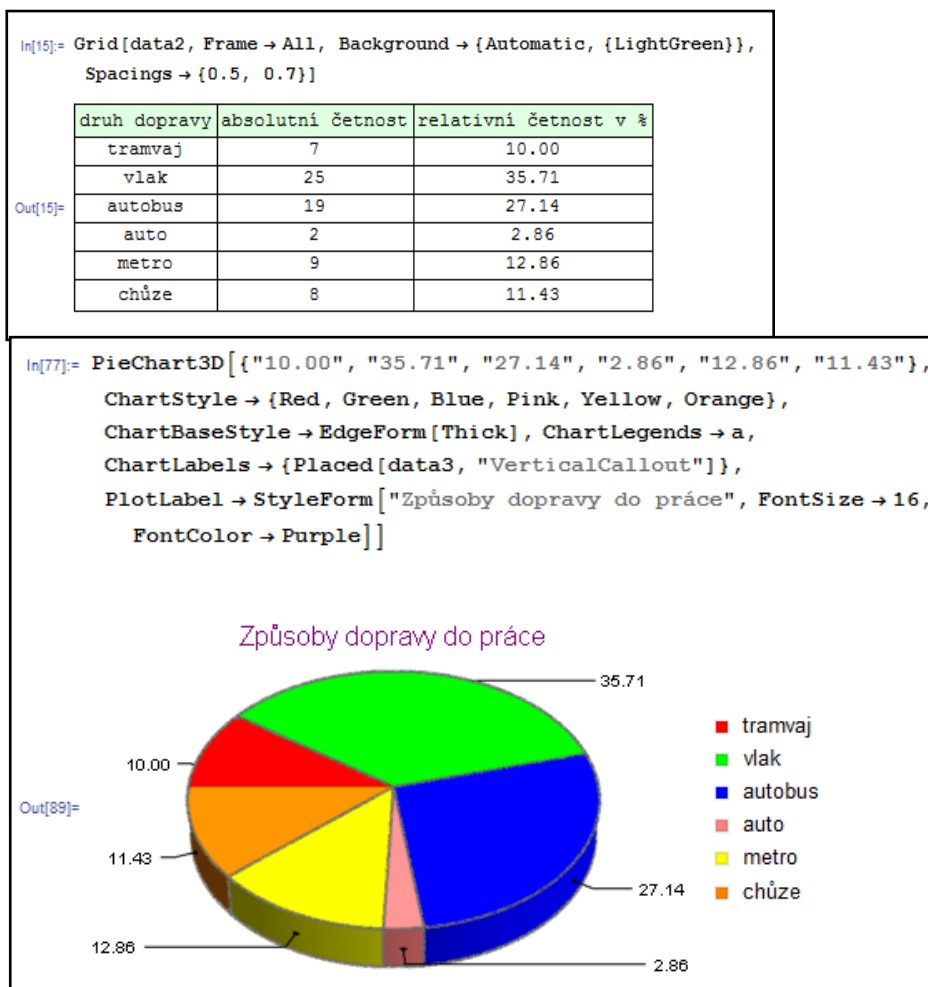
Obrázek 25 Histogram k příkladu č. 7

zdroj: vlastní

4.6 Výšečový graf

Výšečový graf známí také jako koláčový graf vyjadřuje relativní četnosti jednotlivých hodnot statistického znaku, neboli kolika procenty se jednotlivé hodnoty podílejí na celku. Graf je rozdělen na jednotlivé výšeče, jejichž velikost vyjadřuje míru četnosti. Čím je velikost výšeče větší, tím je hodnota statistického znaku zastoupena více. Pro přehlednost jednotlivé výšeče barevně odlišujeme.

Výšečový graf vidíme na obrázku 26. Tento graf vyjadřuje, kolik procent z celkem 70 zaměstnanců využívá různé způsoby dopravy do práce. Podrobné údaje vidíme na obrázku 24 v tabulce. K zjištěným údajům jsme si museli ještě vypočítat relativní četnosti u jednotlivých druhů doprav. Program Mathematica nabízí vytvoření i 3D obrazců, proto jsme zde použili příkaz *PieChart3D*. Pomocí příkazu *ChartStyle* jsme si určili barvu jednotlivých výšečí, přidali jsme ke grafu legendu pomocí *ChartLegends* (písmeno a reprezentuje vybraný 1. sloupec z tabulky) a příkaz *ChartLabels* nám přiřadil popisky k jednotlivým výšečím grafu.



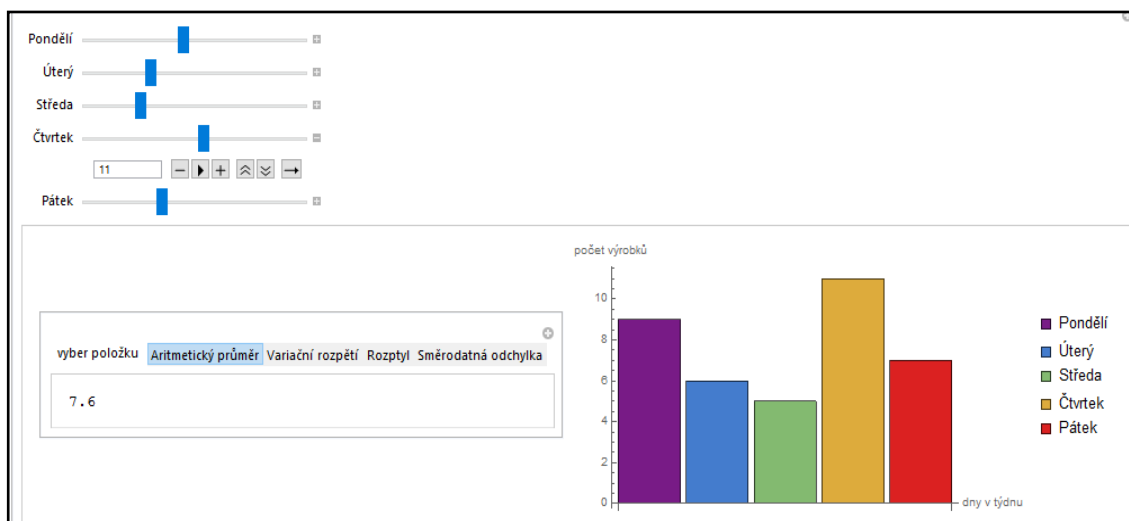
Obrázek 26 Příklad výšečového grafu

zdroj:vlastní

5 PRAKTICKÉ VYUŽITÍ SOFTWARE MATHEMATICA

Software Mathematica nabízí možnost vytváření interaktivních příkladů, které obsahují ovládací prvky jako nastavení vstupní hodnoty, výběr z nabídky položek, zaškrtnutí políčka nebo číselnou osu s posuvníkem. Takovýto příklad můžeme vytvořit pomocí příkazu *Manipulate*. Výstup příkazu *Manipulate* může mít jak číselnou podobu tak i grafickou (měníme parametry grafu). Použití příkazu *Manipulate* si ukážeme na jednoduchém příkladě.

Firma vyrábí produkt a potřebuje analyzovat svá data o počtu vyrobených produktů za týden. K analýze dat slouží nástroj vytvořený v programu Mathematica, který vidíme na obrázku 27. Tento nástroj se skládá ze tří částí. U horního okraje vidíme pět posuvných číselných os, na kterých si můžeme nastavit číselné hodnoty. Firma zde zadává denní počet vyrobených výrobků. Pod tím vidíme rámeček, ve kterém se zobrazují výsledky vybraných charakteristik souboru. Máme zde na výběr aktuálně čtyři charakteristiky, záměrně jsme vybrali pouze ty, které mají pro tento typ dat význam. Vedle rámečku s výpočty se nachází sloupcový graf, jehož výšky jednotlivých sloupců se mění podle toho, jaké hodnoty zadáme na číselných osách. Graf doplňuje výpočty charakteristik tak, aby firma měla kompletní přehled.



Obrázek 27 Nástroj pro výpočet charakteristik

zdroj: vlastní

Aby tento model pro výpočet charakteristik fungoval, museli jsme vytvořit v programu Mathematica příkaz, který vidíme na obrázku 28 (následující strana). Nejprve jsme si definovali jednotlivé vzorce pro výpočty, které vidíme na obrázku jako `In[1]` až `In[4]`. Potom jsme sestavili samotný příkaz, kde jsme použili *Manipulate*. Příkaz *Manipulate* se skládá z několika částí, první píšeme modelovanou funkci, podle

kteřé proměnné je modelováno, minimální a maximální hodnota proměnné a poslední je určen o kolik se mají jednotlivé hodnoty zvětšovat či zmenšovat. Tento příkaz je velmi variabilní, můžete přidávat libovolný počet posuvných číselných os s odlišným rozsahem, můžete zadat nespočet různých vzorců a výpočtů nebo můžete přidat celou škálu grafů, které software Mathematica nabízí.

```

In[1]:= prumer[D1_, D2_, D3_, D4_, D5_] = Mean[{D1, D2, D3, D4, D5}] // N;

In[2]:= vrozpeti[D1_, D2_, D3_, D4_, D5_] = Max[{D1, D2, D3, D4, D5}] - Min[{D1, D2, D3, D4, D5}];

In[3]:= rozptyl[D1_, D2_, D3_, D4_, D5_] = Variance[{D1, D2, D3, D4, D5}] // N;

In[4]:= sodchylka[D1_, D2_, D3_, D4_, D5_] = Sqrt[Variance[{D1, D2, D3, D4, D5}]] // N;

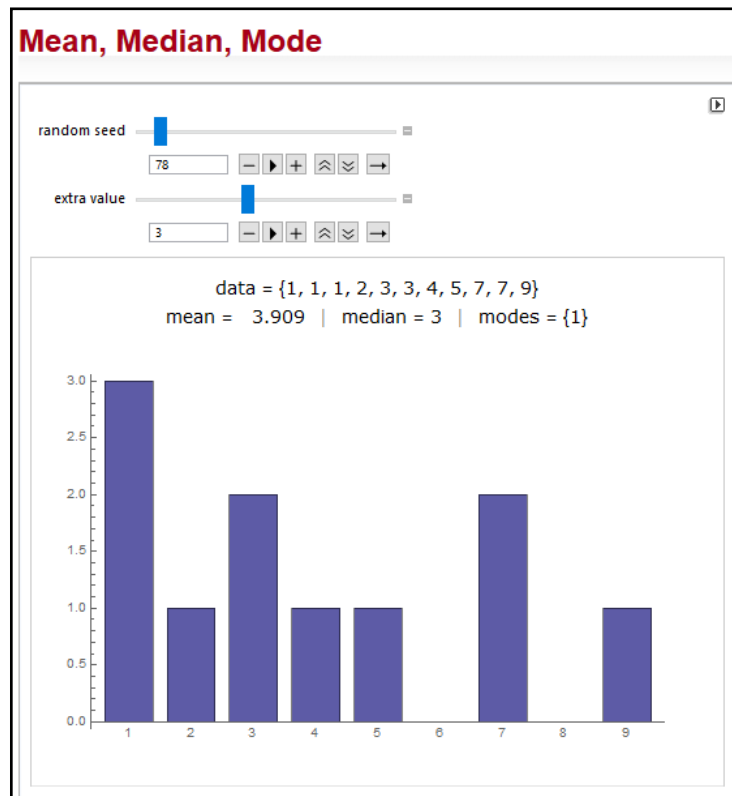
In[5]:= Manipulate[
  Grid[
    {
      {Manipulate[dny[D1, D2, D3, D4, D5],
        {{dny, prumer, "vyber položku"}, {prumer -> "Aritmetický průměr",
          vrozpeti -> "Variální rozpětí", rozptyl -> "Rozptyl", sodchylka -> "Směrodatná odchylka"}},
      BarChart[{D1, D2, D3, D4, D5}, ChartStyle -> "Rainbow",
        AxesLabel -> {"dny v týdnu", "počet výrobků"}, ImageSize -> 400,
        ChartLegends -> {"Pondělí", "Úterý", "Středa", "Čtvrtek", "Pátek"}]},
      {{D1, 0, "Pondělí"}, 0, 20, 1}, {{D2, 0, "Úterý"}, 0, 20, 1}, {{D3, 0, "Středa"}, 0, 20, 1},
      {{D4, 0, "Čtvrtek"}, 0, 20, 1}, {{D5, 0, "Pátek"}, 0, 20, 1}
    }
  ]

```

Obrázek 28 Příkaz k nástroji pro výpočet charakteristik

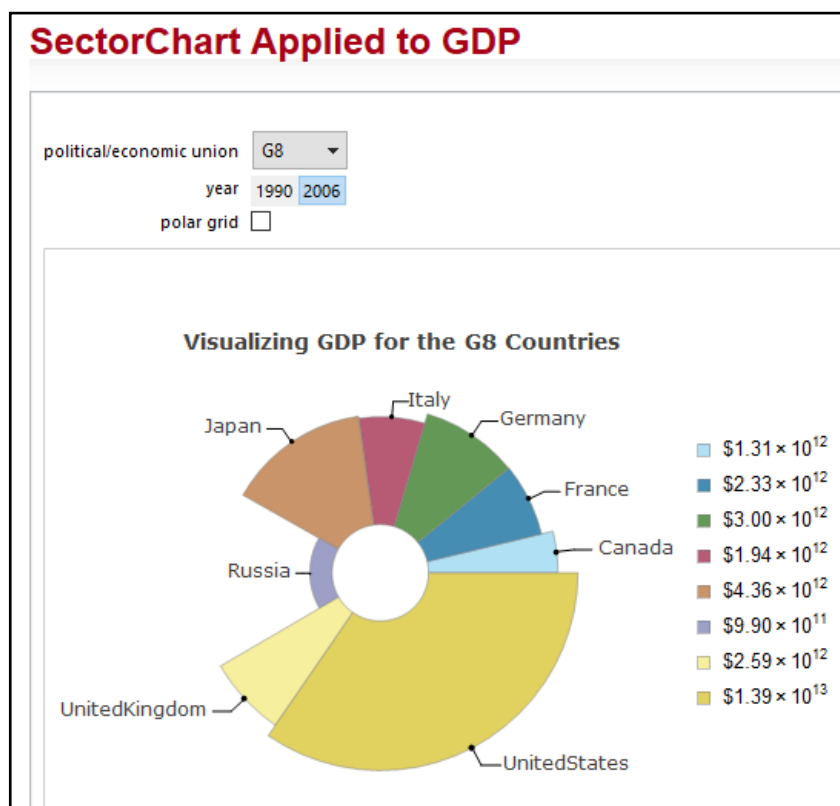
zdroj: vlastní

Další ukázky projektů a příkladů můžeme najít na stránkách <http://demonstrations.wolfram.com/index.html>. Na těchto stránkách najdeme nepřehledné množství vytvořených projektů v programu Mathematica, projekty jsou seřazeny do různých kategorií, např.: Mathematics, Computation, Out Word, Physical Science, Engineering and Technology atd. V kategorii Mathematics je celá sekce věnována statistice. Sekce statistika je dále členěna na 4 kategorie a jsou to: High School Statistics, Data Analysis, Probability a Random Processes. Pro ukázkou si představíme dva projekty. První vidíme na obrázku 29. V tomto projektu zjišťujeme, jak se změní průměr, medián a modus, pokud do seznamu dat přidáme další hodnotu. Nejprve zde vybereme náhodný seznam dat (random seed) a k tomuto seznamu můžeme libovolně přidat další hodnotu (extra value). Je zde umístěn také sloupcový graf, který vyjadřuje četnost jednotlivých hodnot. Druhý projekt (obrázek 30) obsahuje statistiku hrubého domácího produktu GDP v různých politických či ekonomických společenství (záložka political/economic union), můžeme zobrazit data za rok 1990 nebo 2006. Poslední volba se týká zobrazení mřížky v grafu (polar grid).



Obrázek 29 Ukázka projektu-výpočet průměru, mediánu a modu

zdroj: <http://demonstrations.wolfram.com/MeanMedianMode/>



Obrázek 30 Ukázka projektu-statistika hrubého domácího produktu

zdroj: <http://demonstrations.wolfram.com/SectorChartAppliedToGDP/>

ZÁVĚR

Cílem této práce bylo představit program Mathematica a ukázat jak se dá tento program využít pro výpočty z prostředí statistiky, jaké typy příkladů ze statistiky lze pomocí programu řešit a jak jinak můžeme program ve statistice využít. Tento cíl byl během tvorby práce splněn. V 1. kapitole jsme se zabývali samotným programem Mathematica, představili jsme si jeho prostředí, uživatele, programovací jazyk, vznik programu. V následující 2. kapitole jsme se seznámili s pojmem statistika, s historií této vědy a vysvětlili jsme si některé základní pojmy. Celá 3. kapitola je věnována charakteristikám statistického souboru, byli to charakteristiky polohy, variability, šikmosti a špičatosti. U každé charakteristiky najdeme několik příkladů, které doplňují toto téma. Ve 4. kapitole jsme řešili téma rozdělení četností a jejich grafické vyjádření. Popsali jsme několik typů grafů a ukázali jsme si, jak vypadají. V poslední 5. kapitole jsme představili dynamický nástroj pro výpočet charakteristik souboru.

Tato práce poskytuje jasný doklad o tom, že software Mathematica je dostatečně způsobilý nástroj pro řešení příkladů ze statistiky. V této práci jsme řešili příklady převážně z oblasti charakteristik statistického souboru. Pro tuto oblast jsou v Mathematica předdefinované funkce a příkazy, které je možno použít, jako například *Mean*, *GeometricMean*, *Median*, *MeanDeviation*, *Variance*, *Kurtosis*. Pokud potřebný příkaz Mathematica nenabízí, uživatel si může nadefinovat vlastní příkazy a funkce. Vlastní funkce byla využita při výpočtu váženého aritmetického průměru, váženého harmonického průměru, variačního rozpětí, směrodatné odchylky a variačního koeficientu.

Další využití má program Mathematica při tvorbě grafů. Grafy jsou ve statistice velmi důležité, poskytují rychlé a názorné informace o datech. V programu Mathematica můžeme vytvořit nepřeberné množství grafů i ve formátu 3D, v této práci jsme použili příkazy *ListPlot*, *ListLinePlot*, *Histogram*, *BarChart* a *PieChart3D*. U každého grafu si můžeme nastavit další možnosti, přidat popisky os a legendy, zvolit různé barvy i styly grafu.

Software Mathematica je univerzální a velmi multifunkční nástroj pro řešení různých typů úloh, jak z matematiky, statistiky, fyziky, chemie. Nabízí možnost jak početních tak i grafických výstupů. Jeho jedinou nevýhodou je absence nástrojů pro ukládání práce v jiném formátu, s jinou příponou souboru.

ANOTACE

Příjmení a jméno autora:	Kristková Markéta
Instituce:	Moravská vysoká škola Olomouc
Název práce v českém jazyce:	Využití software Mathematica ve statistice
Název práce v anglickém jazyce:	Software Mathematica and Statistic
Vedoucí práce:	Mgr. Veronika Říhová, Ph.D.
Počet stran:	59
Rok obhajoby:	2016
Klíčová slova v českém jazyce:	software Mathematica, statistika, průměr, variabilita, šikmost, špičatost, četnosti
Klíčová slova v anglickém jazyce:	software Mathematica, statistics, mean, variability, skewness, kurtosis, frequency

Cílem této bakalářské práce Využití software Mathematica ve statistice je představit počítačový program Mathematica a ukázat jak se dá tento program využít pro výpočty z prostředí statistiky, jaké typy příkladů ze statistiky lze pomocí programu řešit a jak jinak můžeme program ve statistice využít. Tato práce je rozdělena na 5 hlavních kapitol, první dvě kapitoly obsahují teoretické informace o programu Mathematica a základní informace ze statistiky. Zbývá část práce je věnována vybraným oblastem statistiky, které jsou doplněny řešenými příklady.

The aim of the thesis Software Mathematica and Statistics is to describe computer program Mathematica and demonstrate how this computer program can use for calculations from area of statistics, what kind of examples from statistics program can solve and other utilization in statistics. This thesis is divided into five main parts, first two parts include theoretic information about program Mathematica and basic information about statistics. Remaining part focus on select areas of statistics and this part is supplemented by resolved examples.

LITERATURA A PRAMENY

- ARLTOVÁ, Markéta, aj. *Základy statistiky v příkladech*. 1.vyd. Brno: Tribun EU s.r.o., 2014. 192 s. ISBN 978-80-263-0756-3.
- FŇUKAL, Miloš, SMRČKA, David, a KLADIVO Petr. *Software Mathematica pro geografii*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2015. 81 s. ISBN 978-80-244-4472-7.
- FRIEDRICH, Václav. *Mathematica na počítači pro nematematiky*. 1.vyd. Ostrava: Vysoká škola Báňská-Technická univerzita Ostrava, 2013. 268 s. ISBN 978-80-248-3162-6
- HINDLS, Richard, HRONOVÁ, Stanislava a SEGER Jan. *Statistika pro ekonomy*. 5. vyd. Praha: Professional Publishing, 2004. 415 s. ISBN 80-86419-59-2.
- CHRAMCOV, Bronislav. *Základy práce v prostředí Mathematica*. 2.vyd. Zlín: Univerzita Tomáše Bati ve Zlíně, 2006. 128 s. ISBN 80-7318-510-5.
- KLADIVO, Petr. *Základy statistiky*. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, 2013. 64 s. ISBN 978-80-244-3841-2.
- LÖSTER, Tomáš, ŘEZANKOVÁ, Hana a LANGHAMROVÁ Jitka. *Statistické metody a demografie*. 1. vyd. Praha: Vysoká škola ekonomie a managementu, 2009. 291 s. ISBN 978-80-86730-43-1.
- OTIPKA, Petr a ŠMAJSTRLA, Vladislav. *Pravděpodobnost a statistika*. 1. vyd. Ostrava: Vysoká škola Báňská - Technická univerzita Ostrava, 2006. 266 s. ISBN 80-248-1194-4.
- ŘEZANKOVÁ, Hana, a LÖSTER, Tomáš. *Základy statistiky*. 1. vyd. Praha: Oeconomica, 2013. 95 s. ISBN 978-80-245-1957-9.
- ŠALOUNOVÁ, Dana. *Úvod do pravděpodobnosti a statistiky*. 1. vyd. Ostrava: Vysoká škola Báňská-Technická Univerzita Ostrava, 2013. 173 s. ISBN 978-80-248-3067-4.

Online zdroje

czso.cz, *Česká republika od roku 1989 v číslech*. [online]. 2016 [cit. 2016-02-25].
Dostupné z: <<https://www.czso.cz/csu/czso/ceska-republika-od-roku-1989-v-cislech>>

czso.cz, *Česká republika od roku 1989 v číslech*. [online]. 2016 [cit. 2016-04-18].
Dostupné z: <https://www.czso.cz/csu/czso/ceska-republika-od-roku-1989-v-cislech#12>

czso.cz, *Databáze demografických údajů za obce ČR*. [online]. 2016
[cit. 2016-02-25]. Dostupné z:
<<https://www.czso.cz/csu/czso/databaze-demograficky-udaju-za-obce-cr>>

czso.cz, *Přírůstek (úbytek) obyvatelstva přirozenou měnou v krajích ČR v roce 2015*. [online]. 2016 [cit. 2016-02-25]. Dostupné z:
<<https://www.czso.cz/csu/xm/prirustek-ubytek-obyvatelstva-prirozenou-menou-v-krajich-cr-v-roce-201x>>

demonstrations.wolfram.com, *Mean, Median, Mode*. [online]. 2016
[cit. 2016-04-26]. Dostupné z:
<<http://demonstrations.wolfram.com/MeanMedianMode/>>

demonstrations.wolfram.com, *SectorChart Applied to GDP*. [online]. 2016
[cit. 2016-04-26]. Dostupné z:
< <http://demonstrations.wolfram.com/SectorChartAppliedToGDP/>>

kantaylis.com, *Základní statistické pojmy*. [online]. 2014 [cit. 2016-01-12].
Dostupné z: <http://kantaylis.com/zakladni_statisticke_pojmy.html>

mbenzin.cz, *Ceny benzínu a nafty Praha*. [online]. 2016 [cit. 2016-03-05].
Dostupné z: <http://www.mbenzin.cz/Ceny-benzinu-a-nafty-Praha-M_>

SEZNAM ZKRATEK

ČSU	Český statistický úřad
In	vstupní buňka v programu Mathematica
Out	výstupní buňka v programu Mathematica

SEZNAM OBRÁZKŮ

Obrázek 1 Uvítací okno programu	9
Obrázek 2 Notebook.....	10
Obrázek 3 Příklady použití pravidel.....	12
Obrázek 4 Rozdělení statistických znaků.....	14
Obrázek 5 Řešení příkladu č. 1	19
Obrázek 6 Řešení příkladu č. 2	21
Obrázek 7 Řešení příkladu č. 3	23
Obrázek 8 Řešení příkladu č. 4	25
Obrázek 9 Přepočítání na desetinná čísla	25
Obrázek 10 Řešení příkladu č. 5	27
Obrázek 11 Řešení příkladu č. 6	28
Obrázek 12 Řešení příkladu č. 7	30
Obrázek 13 Řešení příkladu č. 8	32
Obrázek 14 Data k příkladu č. 9.....	35
Obrázek 15 Řešení příkladu č. 9	37
Obrázek 16 Řešení příkladu č. 10	38
Obrázek 17 Grafické znázornění rozložení četností.....	39
Obrázek 18 Grafická interpretace špičatosti	40
Obrázek 19 Řešení příkladu č. 11	41
Obrázek 20 Graf k příkladu č. 11	42
Obrázek 21 Četnost známek z testu	44
Obrázek 22 Četnost prodeje obuvi	45
Obrázek 23 Rozložení platových tříd mezi zaměstnanci	46
Obrázek 24 Četnost dosaženého vzdělání	47
Obrázek 25 Histogram k příkladu č. 7	48
Obrázek 26 Příklad výsečového grafu.....	49
Obrázek 27 Nástroj pro výpočet charakteristik.....	50
Obrázek 28 Příkaz k nástroji pro výpočet charakteristik	51
Obrázek 29 Ukázka projektu-výpočet průměru, mediánu a modu.....	52
Obrázek 30 Ukázka projektu-statistika hrubého domácího produktu	52

SEZNAM TABULEK

Tabulka 1 Počet nehod v letech.....	18
Tabulka 2 Počet obyvatel v Praze	23
Tabulka 3 Přírůstek (úbytek) v krajích v roce 2015	24
Tabulka 4 Počet cizinců s trvalým pobytem	31
Tabulka 5 Ceny pohonných hmot v Praze.....	35
Tabulka 6 Zadání k příkladu č. 10.....	38
Tabulka 7 Počty dětí v mateřských školách	41