ZÁPADOČESKÁ UNIVERZITA V PLZNI

FAKULTA PEDAGOGICKÁ

KATEDRA MATEMATIKY, FYZIKY A TECHNICKÉ VÝCHOVY

STATISTICKÉ GRAFY A JEJICH VYUŽITÍ

BAKALÁŘSKÁ PRÁCE

Veronika Kulhánková

Přírodovědná studia, Matematická studia

Vedoucí práce: RNDr. Václav Kohout

Plzeň, 2014

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně s použitím uvedené literatury a uvedených zdrojů informací.

Plzeň, 5. dubna 2014

……………………………

podpis

PODĚKOVÁNÍ

Ráda bych poděkovala RNDr. Václavu Kohoutovi za podporu a rady při psaní své bakalářské práce.

OBSAH

1 ÚVOD

Důvodem volby tématu mé bakalářské práce byla možnost blíže se informovat o statistických grafech. S nejrůznějšími druhy vizuálního zpracování dat se často setkáváme v běžném životě. Grafy jsou součástí dnešní moderní společnosti. Využívají se k názornější interpretaci dat a je v nich lepší orientace. Málokomu se chce číst několika stránkové dokumenty. Jako příklad bych uvedla vývoj cen. Na grafu je vše přehledně, názorně a dokonce i strategicky prezentováno. Aby grafy cíleného spotřebitele zaujaly, využívají se nejrůznější barevné efekty. Mnoho výrobců rádo zveřejní grafy poklesu cen výrobků, aby oslovili co největší počet lidí. A už jsme u té strategičnosti.

Hlavním cílem této práce je seznámit čtenáře se statistickými grafy.

Práce nejprve vysvětluje data a termíny statistiky. V této úvodní části se věnuji pojmům, jako jsou např. statistika jako vědní obor, statistický soubor, statistický výběr, data, která se dělí do několika skupin stejně jako statistické znaky. Dále se v této kapitole věnuji charakteristikám statistického souboru a sice měrám polohy, variability a tvaru. Pro větší názornost jsem se vše snažila doplnit konkrétními příklady.

V další kapitole už jsem se začala zabývat jednotlivými statistickými grafy. Ty jsou dělené do dvou hlavních skupin – grafy jednorozměrných dat a grafy vícerozměrných dat. Grafy jsou v této části definovány, znázorněny a závěrem diskutovány v různých typech rozdělení. K tvorbě ukázek grafů jsem využila vždy stejná či podobná data, aby bylo vidět, jak který graf data zobrazuje a v čem jsou jeho přednosti či nedostatky.

Poslední kapitola mé práce je zaměřena na využití grafů. Snažila jsem se na dvou různých typech dat grafy porovnávat, diskutovat jejich normalitu a body, které vybočují.

K tvorbě grafů jsem využila prostředí programu Wolfram Mathematica verze 8.0 a Microsoft Excel 2010. V programech je mnoho grafů již předem definováno.

Hlavními zdroji, ze kterých jsem čerpala, byly monografie autorů M. Meloun a J. Militký, dále webové stránky vedoucího mé bakalářské práce RNDr. Václava Kohouta a další internetové zdroje, monografie a již zmíněný program Wolfram Mathematica verze 8.0 a Microsoft Excel 2010.

2 STATISTICKÉ POJMY

2.1 STATISTIKA

Statistika je vědní disciplína zabývající se analýzou dat v přírodních i společenských vědách. Pracuje s daty, které sbírá, zpracovává, vyhodnocuje a interpretuje. Výstupem těchto dat bývají většinou grafy, což závěry činí přehlednějšími.

[Kohout, V.]

2.2 DATA

Data, údaje popisující vlastnosti určitého objektu, jsou hlavním atributem pro statistiku. Tzv. primární data získáváme měřením a pozorováním. Jejich zisk je nákladnější než u dat sekundárních. Příkladem může být získávání těchto dat formou dotazníkového šetření při zjišťování spokojenosti občanů s obchodním pokrytím jejich bydliště.

Dalšími daty jsou tzv. sekundární data. Ta jsou již zpracována a využívají se k dalšímu šetření.

Významnou institucí ČR, která zpracovává data a shromažďuje podrobné statistiky, je Český statistický úřad. Údaje, které získává, pocházejí od jednotlivých ministerstev, která tvoří také své statistické přehledy. Dále je možné získat data z nejrůznějších databází, výzkumných zpráv atp.

[management – marketingu]

2.2.1 Jednorozměrná data

Jednorozměrná data se vyskytují v náhodném skaláru (čísle).

Experimentální data jsou často v malém počtu, asymetrického rozdělení, různého rozptylu a porušují základní předpoklady, které jsou kladeny na výběr.

Vyšetřují se statistické zvláštnosti dat jako např. podezřelé hodnoty, jejichž zobrazení je výrazně viditelné v grafu. Data se většinou vztahují k normálnímu rozdělení a zjišťují se případné odchylky od tohoto typu rozdělení.

[Meloun, M., 1998]

2.2.2 Vícerozměrná data

Vícerozměrná data jsou obsažena v náhodném vektoru. Všechny složky tohoto vektoru jsou neovlivnitelné.

„*Na základě provedených analýz je k dispozici náhodný výběr velikosti n.*“ [dle Meloun, M., 2012] Tento výběr je složen n-ticí vektorů x_{jT} = (x_j, 1, …, $x_{j,m}$). Tyto n-tice je možné chápat jako souřadnice n bodů v m rozměrném prostoru. Takovýto náhodný výběr je potom vyjádřen maticí typu (n x m) a platí n > m.

[Meloun, M., 2006]

Obrázek 1 - Matice vícerozměrných dat

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_i^T \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,i} & \cdots & x_{1,m} \\ \vdots & & \vdots & & \vdots \\ x_{i,1} & \cdots & x_{i,i} & \cdots & x_{i,m} \\ \vdots & & \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,i} & \cdots & x_{n,m} \end{bmatrix}$$

[dle Meloun, M., 2012]

Mezi příklady vícerozměrných dat patří:

- vyjádření vlastností produktů – potravin, sloučenin atd., pomocí různých analytických metod
- hodnocení spekter pomocí poloh a velikostí píků či plochy absorpčních pásů pro charakterizaci a identifikaci chemických sloučenin
- sledování složení surovin, produktů v závislosti na čase nebo místě výskytu

[Meloun, M., 2006]

2.2.3 Kvalitativní data

Kvalitativní data můžeme nazvat nečíselnými charakteristikami zkoumaného jevu jako např. barva, nálada atp. Někdy je označujeme jako data měkká.

[Management Mania]

Sběr těchto dat je snadnější, protože jsou lépe přístupná ve srovnání s daty kvantitativními. Pomáhají při ověřování zjišťovaných závěrů a mohou dosud platnou teorii potvrdit nebo naopak vyvrátit.

[Štěpáníková, Čermák]

NOMINÁLNÍ DATA

O hodnotách proměnné jsme schopni říci, zda jsou identické či odlišné např. model, výrobce atp.

[Management Mania]

Čísla jsou zde využívána pouze, jako označení proměnných tzn., není možné s nimi, jako s čísly, počítat. Příkladem může být číselné označení pohlaví (1- muž, 2 – žena nebo naopak). Z tohoto příkladu je patrné, že se zavedením těchto čísel není možné využít aritmetických operací.

[Základy statistiky]

ORDINÁLNÍ DATA

O hodnotách proměnné jsme schopni říci, zda jsou identické či odlišné a dovedeme určit pořadí proměnné např. míra spokojenosti atp.

[Management Mania]

Čísla se přiřazují tak, že vyjadřují pořadí podle zadaného kritéria např. seřazení žáků podle tělesné výšky (1 – nejvyšší žák …15 – nejnižší žák) atp. Poskytují informace jen o pořadí měřených objektů, o velikostech rozdílů mezi čísly se nic nedozvíme.

[Základy statistiky]

2.2.4 Kvantitativní data

Kvantitativní data udávají číselné charakteristiky zkoumaného jevu jako např. váha, výška, cena atp. Někdy je označujeme jako data tvrdá.

[Management Mania]

O těchto datech můžeme říci, že nám poskytují více informací, než data kvalitativní, protože není nutné tak dlouhého pozorování jako pro získání dat kvalitativních.

[institutu biostatiky a analýz Masarykovi univerzity]

Pomáhají ověřovat platnost dat kvalitativních a snaží se o přesné a zobecnitelné závěry.

[Štěpáníková, Čermák]

INTERVALOVÁ DATA

O hodnotách proměnné jsme schopni říci, zda jsou identické či odlišné a můžeme rozhodnout, o kolik je jedna hodnota větší než druhá.

[Management Mania]

Čísla vyjadřují kvantitativní míru vlastnosti i velikost rozdílů mezi nimi. Je definovaná jednotka měření, která nemá přirozený nulový bod. Čísla, která získáváme intervalovým měřením, je možné sčítat, odečítat, avšak není možné je násobit a dělit. Pro operace sčítání a odečítání je možné získat i sporné hodnoty.

[Základy statistiky]

POMĚROVÁ DATA

O hodnotách proměnné jsme schopni říci, zda jsou identické či odlišné a můžeme rozhodnout, kolikrát je jedna hodnota větší než druhá.

[Management Mania]

„*Čísla vyjadřují kvantitativní míru vlastnosti i velikosti násobků mezi nimi. Kromě přesně definované jednotky měření, zde existuje i přirozená nula*." [dle Základy statistiky] Je zde možné čísla sčítat, odečítat, ale i násobit a dělit.

[Základy statistiky]

Jednotný název pro data intervalová a poměrová je tzv. Metrická data.

[Management Mania]

Tabulka 1 - Dělení dat podle stupnice

TYP STUPNICE	POUŽITÍ PRO DATA
Nominální stupnice	Je možné rozlišit jednotlivé prvky statistického souboru a zařadit je do tříd
Ordinální stupnice	Navíc: Je možné určit rozdíly ve velikosti jednotlivých prvků a zařadit je do tříd
Intervalová stupnice	Navíc: Je možné stanovit relativní počáteční bod a vůči němu určit velikost prvků
Poměrová stupnice	Navíc: Je možné stanovit absolutní počáteční bod a vůči němu určit velikost prvků

[Kohout, V.]

2.3 STATISTICKÝ SOUBOR

Statistický soubor neboli populace je množina všech prvků, které jsou zájmem statistického výzkumu. Tyto prvky se nazývají statistickými jednotkami.

Většina populací je konečná. Nekonečné mohou být statistické soubory určené znakem, které je možné teoreticky nekonečněkrát opakovat.
[Kohout, V.]

Příkladem konečného statistického souboru může být např. počet prodaných automobilů v roce 2013. Nekonečný statistický soubor je potom třeba pásová výroba.

Podle počtu použitých statistických znaků dělíme statistické soubory:

2.3.1 Jednorozměrný statistický soubor

Jednorozměrný statistický soubor sleduje jeden statistický znak. [Kohout, V.]

Např. Počet prodaných automobilů jisté značky v roce 2013.

2.3.2 Vícerozměrný statistický soubor

Vícerozměrný statistický soubor sleduje dva a více statistických znaků. [Kohout, V.]

Např. počet prodaných automobilů pěti různých značek ve státech EU roce 2013.

2.4 STATISTICKÝ VÝBĚR

Informace o souboru získáváme statistickým výběrem. Nepoužíváme všechny dostupné statistické jednotky a tyto statistické jednotky jsou vybírané záměrně či náhodně. Množství statistických jednotek nazýváme rozsah statistického výběru, který dělíme podle velikosti na malý statistický výběr (do 50) a velký statistický výběr (stovky až statisíce).

Podle způsobu, jakým byl statistický výběr proveden, dělíme na statistický výběr:

- Záměrný (opírá se o expertní stanoviska k vytvoření representativního výběru. Často záleží na subjektu experta, vybíráme jednotky něčím typické)
- Náhodný (dochází ke statistickému výběru statistických jednotek ze statistického souboru náhodně bez závislosti na subjektech.)
    - prostý (často je využívána metoda losování)
    - oblastní (soubor rozdělíme na „podsoubory“ a z nich provádíme statistický výběr)
    - mechanický nebo také systematický (vybíráme vždy několikátou jednotku co do pořadí při realizaci statistického výběru)
    - skupinový (v případě statistických souborů s četností v řádech tisíců až statisíců, rozdělujeme statistický soubor na menší celky o stejné velikosti a s co nejmenší variabilitou.)
    - vícestupňový (provádí se, pokud existuje hierarchický popis celého statistického souboru)

[Kohout, V.]

2.5 STATISTICKÝ ZNAK

Statistické jednotky mohou být dány výčtem nebo tzv. identifikačními znaky, což jsou statistické jednotky se společnými vlastnostmi. Umožňují určit, zda jednotka do daného statistického souboru patří či nikoli.

[Kohout, V.]

Dělení statistických znaků:

Podle použitého měřítka (rozlišení podle toho, co vypovídají jednotlivé statistické znaky) :

2.5.1 Kvalitativní statistické znaky

Kvalitativní statistické znaky se většinou vyjadřují textem. [Kohout, V.]

NOMINÁLNÍ STATISTICKÉ ZNAKY (nominální měřítko)

Umožňuje rozlišit jednotlivé hodnoty. U tohoto typu měřítka je možné určit počet kategorií, které jsme použili. Tyto kategorie jsou navzájem neslučitelné. Pokud rozlišujeme právě dvě kategorie, jedná se o dichotomický statistický znak. U více kategorií jde o statistický znak polytomický.

Platí zde pouze vztahy $x_a = x_b$ a $x_a = x_b$, měřených na dvou objektech A a B.

Příkladem je dělení studentů podle pohlaví - student, studentka.

[Hendl, J., 2009]

ORDINÁLNÍ STATISTICKÉ ZNAKY (ordinální měřítko)

Statistické znaky ordinálního měřítka jsou definovány stejně jako znaky nominálního měřítka rozšířené o přesné uspořádání podle jejich velikosti.

[Kohout, V.]

Platí zde navíc vztahy $x_a > x_b$, $x_a < x_b$, měřených na dvou objektech A a B.

[Hendl, J., 2009]

Příkladem může být školní prospěch studentů (podprůměrný, průměrný, nadprůměrný). Je ale, že i když jsou vzdálenosti mezi jednotlivými známkami stejné, dochází k subjektivnímu hodnocení a tato vzdálenost se pak liší. Tzn., pedagogovi se lépe hodnotí známkou chvalitebně místo výborně ve srovnání s dostatečným a nedostatečný, kde toto rozpětí bývá výrazně větší.

2.5.2 Kvantitativní statistické znaky

Kvantitativní statistické znaky je možné vyjádřit číslem. [Kohout, V.]

INTERVALOVÉ STATISTICKÉ ZNAKY (intervalové měřítko)

Intervalové měřítko má stejné vlastnosti jako měřítko ordinální, ale navíc lze údaje uvnitř systému sčítat a odečítat, protože jsou dány nějakou jednotkou měření.

Může zde být definována velikost rozdílu, takže objekt A je rozdílný od objektu B o $x_a - x_b$.

U intervalového měřítka je možné zvolit počátek. Příkladem intervalového měřítka je Celsiova stupnice. Intervalovým znakem by byla teplota měřená na této stupnici.

[Přehled statistických metod]

POMĚROVÉ STATISTICKÉ ZNAKY (poměrové měřítko)

Toto měřítko má stejné vlastnosti jako měřítko intervalové rozšířené o existenci absolutního nulového bodu.

Protože je zde navíc definována absolutní nula platí, že A je x_a / x_b větší než B pro $x_a > x_b$ a $x_b = 0$.

[Hendl, J., 2009]

Údaje jsou vztaženy k pevně daným jednotkám. Příkladem poměrového měřítka je Kelvinova stupnice. Poměrovým znakem by byla teplota měřená na této stupnici.

[Kohout, V.]

Intervalové a poměrové měřítko souhrnně označujeme jako METRICKÁ MĚŘÍTKA, potom mluvíme o metrických statistických znacích. Rozšíříme-li metrická měřítka ještě o měřítko ordinální, nazveme tuto skupinu INTENZIVNÍMI MĚŘÍTKY. Znaky pak nazýváme intenzivními statistickými znaky.

[Hendl, J., 2009]

Obrázek 2 - Dělení statistických znaků

3 CHARAKTERISTIKY STATISTICKÉHO SOUBORU

Mějme statistický soubor X, potom libovolný prvek tohoto statistického souboru je x_i.

Definujme dále využívané pojmy:

ABSOLUTNÍ ČETNOST prvku x_i je definována celkovým počtem prvků x_i ve statistickém výběru.

RELATIVNÍ ČETNOST prvku x_i je definována jako podíl absolutní četnost k celkovému počtu prvků x_i ve statistickém výběru.

KUMULATIVNÍ ABSOLUTNÍ ČETNOST prvku x_i je definována jako součet celkových absolutních četností prvků. Tyto prvky jsou menší nebo rovny prvku x_i.

KUMULATIVNÍ RELATIVNÍ ČETNOST prvku x_i je definována jako součet celkových relativních četností prvků. Tyto prvky jsou menší nebo rovny prvku x_i.

[Kohout, V.]

Pro výpočet jednotlivých konkrétních charakteristiky využijeme následující data v Tabulce 2.

Tabulka 2 - Měsíční nájem 15nájemníků

Nájem (Kč)	6500	6900	7400	7800	8100	8200	8600	9100	9200	9300	9500
Četnost	2	2	1	1	1	1	3	1	1	1	1
Relativní četnost	0,13	0,13	0,067	0,067	0,067	0,067	0,2	0,067	0,067	0,067	0,067

3.1 MÍRY POLOHY

Míry polohy jsou číselnými charakteristikami, které pomáhají data vhodně shromažďovat a určují polohu míst, kde se data koncentrují nejvíce.

[Kohout, V.]

3.1.1 Aritmetický průměr ($\overline{X}$)

Aritmetický průměr je využíván u kvantitativních statistických znaků. Je citlivý na odlehlé hodnoty.

[Kohout, V.]

Výpočet pro n hodnot x_1, x_2, x_3, … x_n :

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

[dle Kohout, V.]

Aritmetický průměr pro data uvedená v Tabulce 2:

$$\overline{X} = \frac{121200}{15} = 8080$$

Pokud jsou data uvedena včetně absolutních četností, jedná se o **vážený aritmetický průměr**. Data jsou zde rozdělena do k - skupin po n_k prvcích.

[Kohout, V.]

$$\overline{X} = \frac{\sum_{i=1}^{k} n_i x_i}{n}$$

[dle Kohout, V.]

Pro výpočet váženého aritmetického průměru rozdělíme nejprve data do intervalů a vypočteme středy těchto intervalů.

Aritmetický průměr pro data uvedená v Tabulce 2:

Tabulka 3 - Rozdělení na intervaly

Interval	Četnost	Střed intervalu
6500 – 7000	4	6750
7000 – 7500	1	7250
7500 – 8000	1	7750
8000 – 8500	2	8250
8500 – 9000	3	8750
9000 – 9500	4	9250

$$\overline{X} = \frac{121750}{15} = 8117$$

Při výpočtu váženého aritmetického průměru je důležité zvolit správné rozpětí intervalů. Abychom dosáhli rozdělení intervalů na stejné velikosti, využívá se tzv. Sturgesovo pravidlo (k ~ 1 + 3,3 . $\log n$).

[Kohout, V.]

V našem příkladu dosahuje hodnoty n = 4,88 (tedy 5).

3.1.2 Modus (Mo)

Modus je nejčastěji se vyskytující hodnota. Je možné ho zjišťovat u libovolných statistických znaků. Pokud existují dvě hodnoty se stejnými nejčastějšími četnostmi, uvádí se jako modus obě. Takovýto statistický znak potom nazýváme bimodálním. Pokud je jich více, modus se neurčuje.

[Kohout, V.]

Modus pro data uvedená v Tabulce 2:

Mo = 8600

3.1.3 Medián (Me)

Definujme nejprve kvantily. Výběr je nutné nejprve vzestupně seřadit a potom pro daný p% kvantil určíme pořadové číslo jednotky n_p (n je počet prvků ve statistickém výběru). [Kohout, V.]

$$n.\frac{p}{100} < n_p < n.\frac{p}{100} + 1$$

[dle Kohout, V.]

Jako medián potom označujeme hodnotu p = 50%. Dělí tedy statistický výběr na dvě části o stejné velikosti.

Používá se u statistických znaků, které je možné seřadit. Jde tedy o statistické znaky kvantitativní a ordinální. Vypovídá o symetrii.

Příkladem je výpočet mediánového platu kvantitativních dat. Mediánový plat je ve srovnání s platem průměrným spravedlivější.

Mezi další kvantily zařazujeme:

Dolní kvartil $x_{0,25}$ je 25% kvantilem.

Horní kvartil $x_{0,75}$ je 75% kvantilem.

Kvartily slouží k přehlednému zobrazení dat v krabicovém grafu. Dělí data na vnitřní, vnější a vybočující data neboli outliers.

[Kohout, V.]

Medián pro data uvedená v Tabulce 2:

Me = 8200

3.1.4 Geometrický průměr (G)

Geometrický průměr se využívá jen v případě kladných prvků statistického výběru a na poměrové statistické znaky. Určuje průměrnou změnu velikosti za předpokladu, že je tato změna konstantní.

[Kohout, V.]

Využívá se také ve finanční matematice při zjišťování průměrného úroku.

Geometrický průměr G se vypočte jako n – tá odmocnina součinu prvků x_i.

$$G = \sqrt[n]{x_1.x_2.\dots.x_n}$$

[dle Kohout, V.]

Geometrický průměr pro data uvedená v Tabulce 2:

$$G = \sqrt[15]{3{,}63.10^{58}} = 8017$$

3.1.5 Harmonický průměr (H)

Harmonický průměr se vypočte jako podíl počtu hodnot n a součtu převrácených hodnot statistického výběru. [Kohout, V.]

Příkladem využití může být výpočet průměrné rychlosti vozidla.

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

[dle Kohout, V.]

Harmonický průměr pro data uvedená v Tabulce 2:

$$H = \frac{15}{0{,}00189} = 7936{,}5$$

Pouze míry polohy nám nepostačují k přesnému popisu statistického výběru. Velké množství dat má stejné nebo velmi podobné hodnoty jednotlivých parametrů měr polohy. Přesto ale mohou být odlišné. Proto pro popis výběru používáme další charakteristiky. Jsou jimi míry variability a míry tvaru.

[Kohout, V.]

3.2 MÍRY VARIABILITY

3.2.1 Variační rozpětí (R)

Variační rozpětí R je rozdílem největší a nejmenší hodnoty statistického výběru. Nevýhodou však je, že variační rozpětí, je velmi citlivé na hodnoty, které jsou odlehlé, a dá se použít jen u malých statistických výběrů, protože je nestabilní vzhledem k počtu členů statistického výběru.

[Kohout, V.]

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

[dle Kohout, V.]

Variační rozpětí pro data uvedená v Tabulce 2:

R = 9500 – 6500 = 3000

3.2.2 Rozptyl (s^2) a Směrodatná odchylka (s)

Rozptyl se využívá při měření velikosti čtverců odchylek jednotlivých hodnot statistického výběru od průměru. [Kohout, V.]

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2$$

[dle Kohout, V.]

Rozptyl pro data uvedená v Tabulce 2:

$$s^2 = \frac{14944000}{14} = 1067428{,}6$$

Odmocninou z rozptylu je směrodatná odchylka (s). [Kohout, V.]

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

[dle Kohout, V.]

Směrodatná odchylka pro data uvedená v Tabulce 2:

$$s = \sqrt{1067428{,}6} = 1033{,}2$$

3.2.3 Variační koeficient (VK)

Variační koeficient se využívá při porovnání variability různých statistických znaků ve statistickém výběru nebo mezi statistickými výběry. [Kohout, V.]

Využívá se u kladných statistických znaků.

Při výpočtu variačního koeficientu se průměrná hodnota nesmí rovnat 0.

$$v = \frac{s}{\overline{x}} . 100\%$$

[dle Kohout, V.]

Variační koeficient pro data uvedená v Tabulce 2:

$$v = \frac{1033{,}2}{8080} . 100\% = 13\%$$

3.3 MÍRY TVARU

Pomocí míry tvaru hodnotíme tvar rozdělení a podobnost s rozdělením normálním. [Meloun, M., 2006]

3.3.1 Šikmost (a)

Šikmost je míra tvaru, která znázorňuje tvar rozdělení, jeho šikmost resp. jeho souměrnost či nesouměrnost.

Pro a = 0 jedná se o rozdělení symetrické.

Pro a > 0 prodlužuje se pravý chvost rozdělení.

Pro a < 0 dochází k prodlužování levého chvostu rozdělení.

[Hendl, J., 2009]

$$a = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{n.s^3}$$

[dle Kohout, V.]

Šikmost pro data uvedená v Tabulce 2:

$$a = \frac{-3\,942\,240\,000}{1{,}654.10^{10}} = -0{,}238$$

3.3.2 Špičatost (b)

Koeficient špičatosti ukazuje odchylku špičatosti zkoumaného rozdělení od normálního rozdělení.

Pro b = 0 jde o normální rozdělení.

Pro b > 0 je graf empirické hustoty více plochý.

Pro b < 0 jsou grafy špičatější.

[Hendl, J., 2009]

Následující vzorec je definován v normálním rozdělení. Proto se odečítá -3, aby toto rozdělení zůstalo zachováno.

$$b = \frac{\sum_{i'1}^{n}(x_i - \overline{x})^4}{n.s^4} - 3$$

[dle Kohout, V.]

Špičatost pro data uvedená v Tabulce 2:

$$b = \frac{2,57.10^{13}}{1,709.10^{13}} - 3 = -1,496$$

Po výpočtu míry tvaru jsme o datech v Tabulce 2 zjistili, že šikmost i špičatost nabývají záporných hodnot. Dochází tedy k prodlužování levého chvostu a graf bude špičatější.

4 GRAFY

Grafy nám umožňují estetické vizuální zpracování statistických dat. Slouží k větší přehlednosti, možnosti orientace a studují vztahy mezi jednotlivými statistickými znaky.

Grafy jsou tvořeny v prostředí programu Mathematica 8.0 a Microsoft Excel 2010.

4.1 GRAFY JEDNOROZMĚRNÝCH DAT

K zobrazení jednotlivých grafů je využito dat v normálním rozdělení, které je nejběžněji užívaným.

4.1.1 Graf polosum

Osa x: pořádkové statistiky $x_{(i)}$

Osa y: $Z_{(i)} = 0.5\,(x_{(n+1-i)} + x_{(i)})$

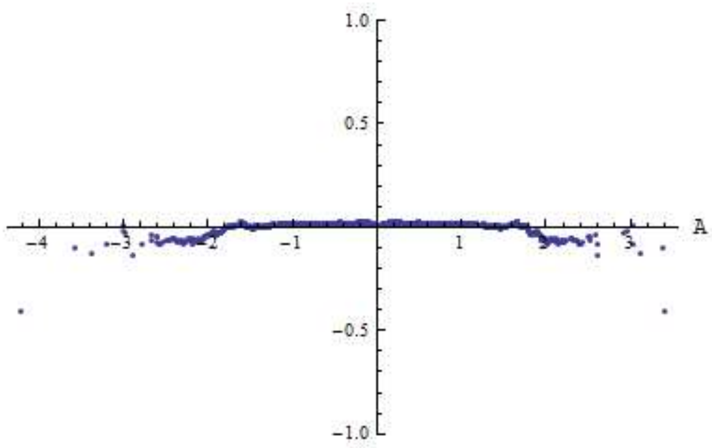
Symetrické rozdělení

Grafem polosum je přímka v horizontální poloze, kterou určuje rovnice y = M, kde M je medián. Typické je detailní měřítko osy y a body kolísají kolem horizontální přímky.

Asymetrické rozdělení

Na rozdíl od symetrického rozdělení není měřítko osy y tak detailní a body nekolísají kolem horizontální přímky. [Meloun, M., 2006]

Graf 1 - Graf polosum v normálním rozdělení



[Kohout, V.]

Z Grafu 1 je patrné, že body kmitají kolem horizontální přímky, ale u obou chvostů je toto kmitání asymetrické a je možné identifikovat vybočující body.

4.1.2 Graf šikmosti

Osa x: $u_{P_i}^2/2$ pro $P_i =$ i/(n + 1)

Osa y: $Z_{(i)} = 0.5\ (x_{(n+1-i)} + x_{(i)})$

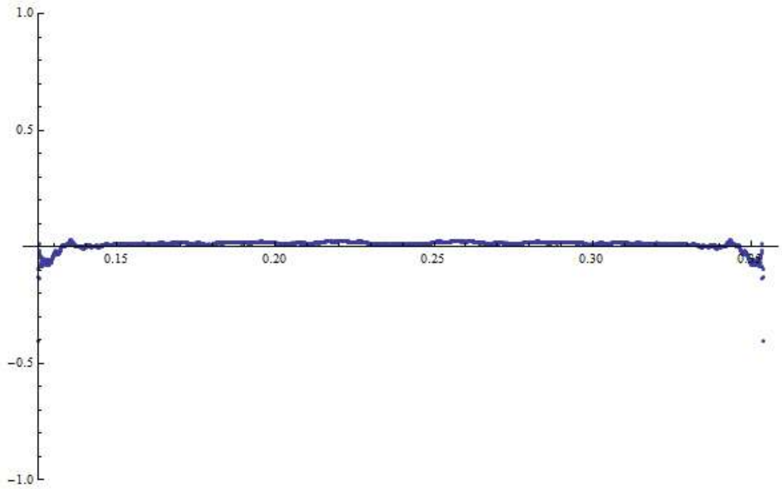
Symetrické rozdělení

Graf v tomto případě inklinuje k přímkové závislosti s nulovým úsekem a jednotkovou směrnicí.

Asymetrické rozdělení

Body na přímce neleží a mají jinou směrnici. [Meloun, M., 2006]

Graf 2 - Graf šikmosti v normálním rozdělení



[Kohout, V.]

Graf 2 ukazuje stejně jako graf předchozí na symetrické rozdělení, ale v oblasti chvostů je graf asymetrický.

4.1.3 Graf symetrie

Osa x: $M - x_{(i)}$

Osa y: $x_{(n+1-i)} - M$

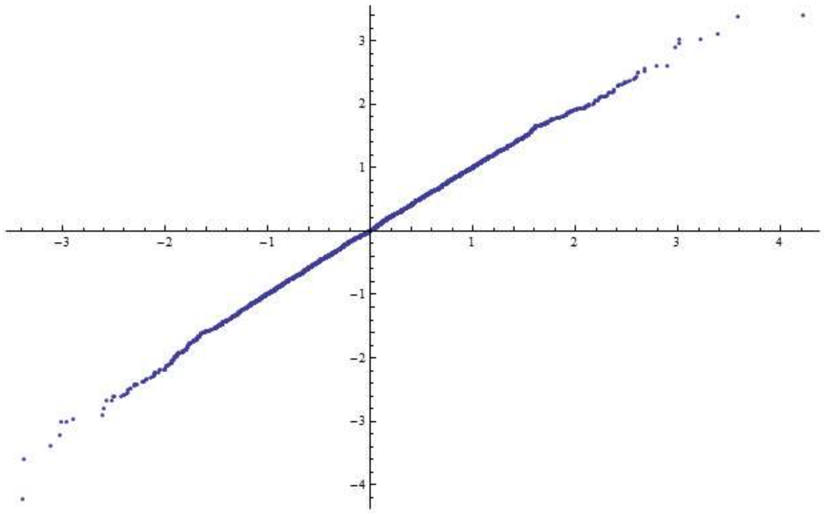
Symetrické rozdělení

Toto rozdělení je určeno přímkou y = M.

Asymetrické rozdělení

V tomto rozdělení postrádá přímka nulovou směrnici a body jsou uspořádané v trendu křivky. [Meloun, M., 2006]

Graf 3 - Graf symetrie v normálním rozdělení



[Kohout, V.]

Graf 3 opět ukazuje na symetrické rozdělení dat s výjimkou oblasti chvostů.

4.1.4 Graf špičatosti

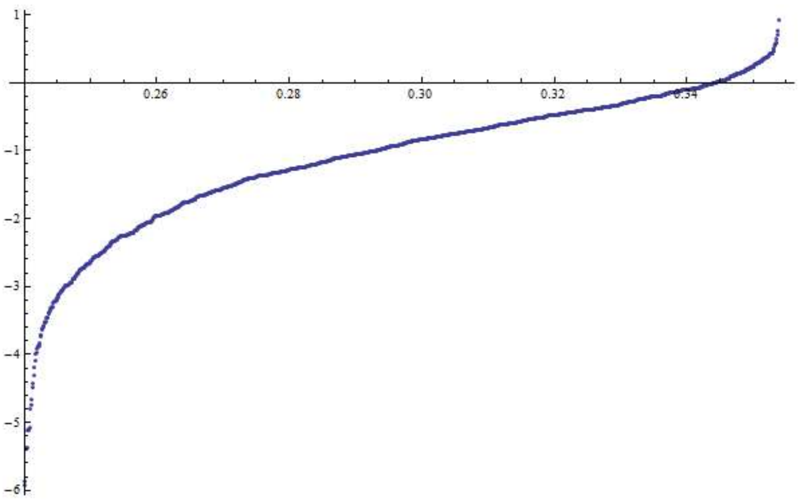
Osa x: $u_{P_i}^2/2$ pro $P_i = i / (n + 1)$

Osa y: ln ($x_{(n+1-i)}$/ $-2u_{P_i}$)

Pro normální rozdělení je grafem přímka v horizontální poloze a body leží většinou na této přímce. Pokud body nejsou náhodné, je hodnota směrnice rovna parametru špičatosti.

[Meloun, M., 2006]

Graf 4 - Graf špičatosti v normálním rozdělení



[Kohout, V.]

Stejně jako u předcházejících grafů se jedná o symetrické rozdělení s výjimkou oblasti chvostů, které se jeví spíše asymetricky.

4.1.5 Kvantilový graf

Osa x: pořadová pravděpodobnost P_i

Osa y: pořádková statistika $x_{(i)}$

Výhodou toho grafu je přehledné zobrazení dat a snadnější rozpoznání tvaru rozdělení, které může být symetrické, zešikmené k nižším nebo vyšším hodnotám.

Do tohoto grafu se většinou zakreslují i kvantilové funkce normálního rozdělení pro snadnější porovnání s normálním rozdělením.

[Meloun, M., 2006]

4.1.6 Kvantilově – kvantilový graf (Q – Q)

Osa x: $Q_T(P_i)$

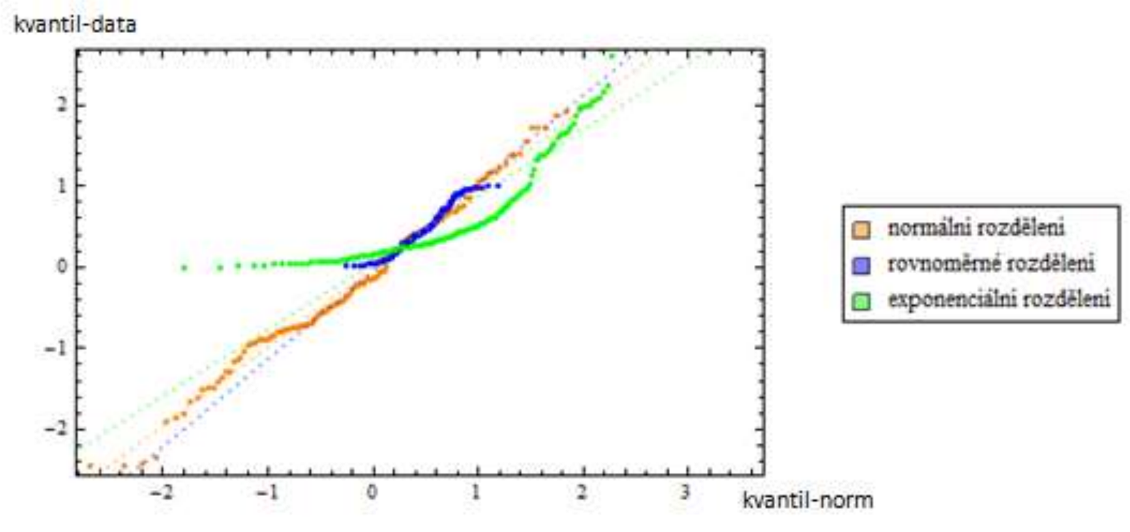
Osa y: $x_{(i)}$

Umožňuje posoudit shodu výběrového rozdělení, které je charakterizováno kvantilovou funkcí $Q_E(P)$ S kvantilovou funkcí zvoleného teoretického zaměření $Q_T(P)$. Jako odhad kvantilové funkce výběru se využívají pořádkové statistiky $x_{(i)}$. Pokud dochází ke shodě výběrového rozdělení s teoreticky zvoleným rozdělením, grafem je přibližně přímka.

Pokud porovnáváme rozdělení výběru s normálním rozdělením s Q – Q grafu, jedná se o rankitový graf.

[Meloun, M., 2006]

Graf 5 - Kvantilově - kvantilový graf



Jak již bylo výše zmíněno, kvantilově – kvantilový graf znázorňuje rozdíly mezi teoretickým rozdělením a rozdělením dat statistického výběru. Tečkovaná čára v grafu naznačuje teoretický průběh rozdělení.

Velké odchylky od normálního rozdělení jsou patrné u dat rovnoměrného a exponenciálního rozdělení především v oblasti chvostů, čímž je tento graf typický. U dat exponenciálního rozdělení je výrazná šikmost a jsou zde patrné odchylky i ve střední části grafu.

4.1.7 Krabicový graf

Osa x: úměrná hodnotám x

Osa y: interval úměrný hodnotě $\sqrt{n}$

Krabicový graf se využívá k částečné sumarizaci dat. Umožňuje znázornění robustního odhadu polohy, mediánu M, posouzení symetrie v okolí kvarilů a u konců rozdělení a dále umožňuje identifikaci odlehlých dat.

Jedná se o graf obdélníkového typu o délce $R_F = F_H - F_D$ se šířkou, která je úměrná hodnotě $\sqrt{n}$.

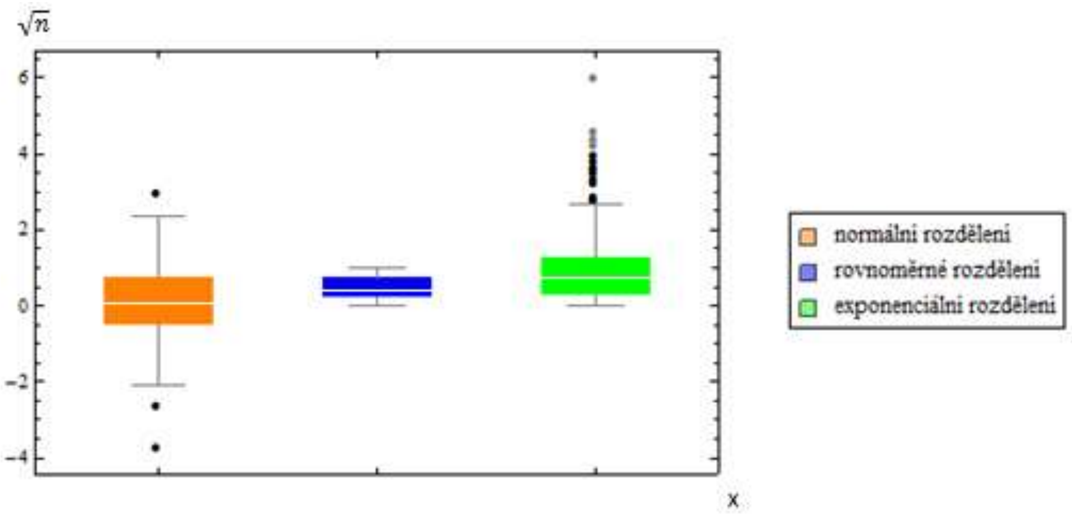
V místě mediánu je vertikální čára a od obou protilehlých stran obdélníku pokračují úsečky, které jsou ukončeny vnitřními hradbami B_H a B_D. Pro tyto hradby platí $B_H = F_H + 1.5\ R_F$,

$B_D = F_D + 1.5\ R_F$. Prvky ležící mimo interval těchto hradeb jsou většinou zakresleny kroužky a jsou to body vybočující tzv. outliers.

Obdobou krabicového grafu je potom graf vrubový krabicový, který navíc umožňuje posouzení variability mediánu.

[Meloun, M., 2006]

Graf 6 - Krabicový graf



Graf 6 zobrazuje data v normálním, rovnoměrném a exponenciálním zobrazení. U normálního rozdělení se vyskytuje malé množství vybočujících bodů (oustliers) a hodnota mediánu je přibližně 0,1. Data rovnoměrného rozdělení jsou lehce asymetrická a nevyskytují se zde žádné vybočující bodu. Naopak u dat exponenciálního rozdělení je výrazné množství vybočujících bodů a výrazná šikmost a asymetrie.

4.1.8 Histogram

Většinou slouží k zobrazení absolutních či relativních četností spojitých znaků.

Osa x: proměnná x

Osa y: úměrná hustotě pravděpodobnosti

Na ose x se nachází jednotlivé třídy, které určují šířky sloupců. Výšky potom znázorňují hustotu pravděpodobností. Čím více máme tříd, tím je graf kvalitnější, podrobnější.

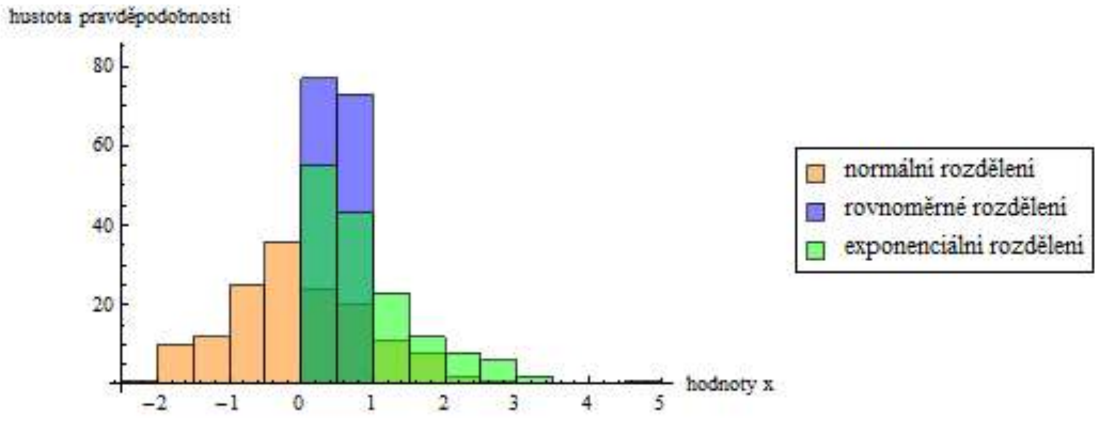
Pro symetrické rozdělení výběru vyčíslujeme počet tříd L podle L = int(2 $\sqrt{n}$). Int(x) je funkce, která je celočíselnou částí čísla x.

[Meloun, M., 2006]

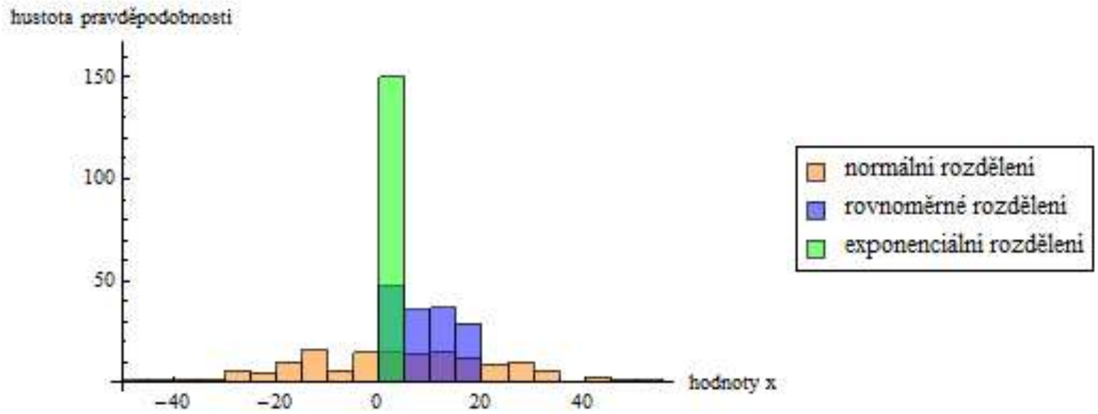
Při předpokladu menšího zešikmení rozdělení statistického výběru je lépe volit stejnou délku Δ x. Při normálním rozdělení je tato hodnota rovna 3.49 s/$n^{1/3}$ (s - směrodatná odchylka). Pokud jsou data přibližně rozdělená je robustní odhad Δx = 2 $F_H - F_D / n^{1/3}$ (F_H, F_D – výběrové kvantily). Při znalosti všech hraničních bodů jednotlivých tříd vypočteme histogram jako (x) = (1/n(x_{j+1}+ - x_j+))* C(x_j+, x_{j+1}+) pro $x_j+ \leq x \leq x_{j+1}+$. C($x_j$+, x_{j+1}+) je funkce, která se rovná celkovému počtu prvků ve statistickém výběru nacházejících se v intervalu $x_j+ \leq x \leq x_{j+1}+$. U jednodušších histogramů se výšky jednotlivých sloupců rovnají absolutním četnostem ve třídách. Takovéto histogramy mají spíše charakterizační funkci.

[Špírková,I., 2011]

Graf 7 - Histogram 1



Graf 8 - Histogram 2



Na Grafech 7 a 8 porovnáváme data v normálním, rovnoměrném a exponenciálním rozdělení. Grafy se liší intervaly, na kterých jsou sestrojeny. Velký rozdíl zaznamenáváme především u rovnoměrného a exponenciálního rozdělení. Na větším intervalu cca (-50,50) se hodnoty dat exponenciálního rozdělení v intervalu (0,5) jeví výrazně vyšší než data rovnoměrného rozdělení. Na menším intervalu cca (-2.5, 5) jsou hodnoty dat nižší než hodnoty dat rovnoměrného rozdělení.

Na první pohled je patrné výrazné zešikmení u dat exponenciálního rozdělení v obou grafech.

Data obojího porovnávaného rozdělení se celkem výrazně liší od pro posouzení uvedených dat normálního rozdělení.

4.1.9 Pravděpodobnostní graf (P – P)

Osa x: P_i

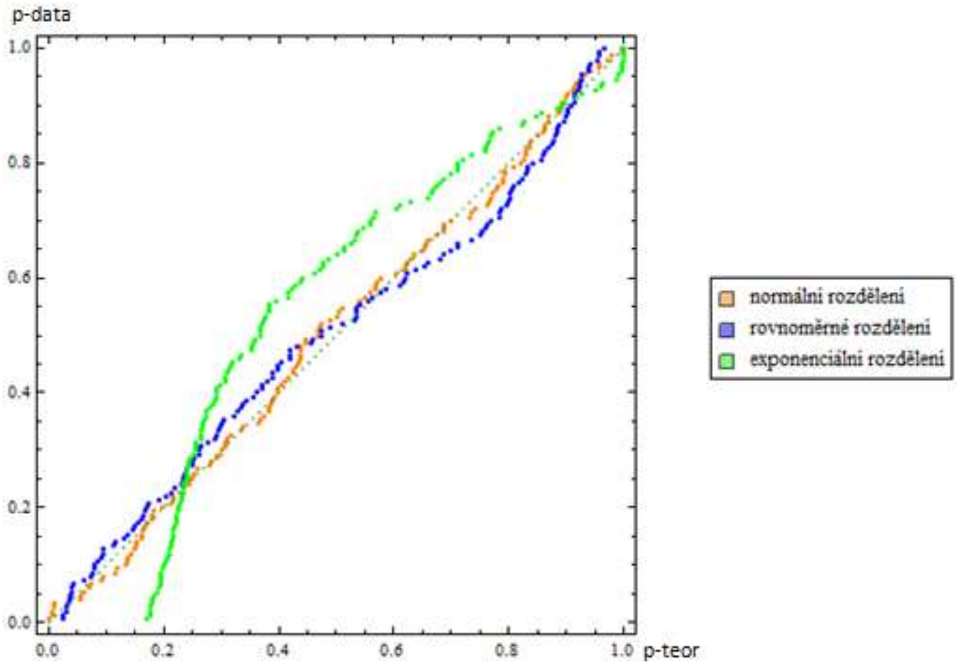
Osa y: $F_T(S_{(i)})$

Pravděpodobnostní grafy jsou jistou alternativou grafů Q – Q. Porovnávají distribuční funkci výběru se standardizovanou distribuční funkcí zvoleného teoretického rozdělení.

U tohoto typu grafu je nutné znát teoretické rozdělení hodnot všech parametrů. Pravděpodobnostní grafy jsou citlivé na odchylky od teoretického rozdělení ve střední části na rozdíl od grafu Q – Q, který je citlivý na tyto odchylky v oblasti konců. Oba grafy se tedy doplňují.

[dle Meloun, M., 2006]

Graf 9 - Pravděpodobnostní graf



Výhodou pravděpodobnostního grafu oproti kvantilově – kvantilovému grafu jsou stejné stupnice.

Na Grafu… porovnáváme opět data rozdělení normálního, rovnoměrného a exponenciálního. Z grafu je patrná výrazná citlivost kolem střední části především u dat rozdělení rovnoměrného a exponenciálního. U exponenciálního je výrazná odlišnost i v oblasti chvostů. Data exponenciálního rozdělení jsou zešikmena. Z grafu můžeme vyčíst i lehkou asymetrii u dat normálního rozdělení, větší potom u dat rozdělení rovnoměrného.

4.1.10 Kruhový graf

Využívá se k ověření hypotézy, zda výběr pochází ze symetrického rozdělení. Potom je grafem polygon blížící se tvarem kružnici. U asymetrického rozdělení je polygonem protáhlý elipsovitý tvar a u rozdělení rovnoměrného jde o elipsovitý tvar podél osy x.

[Meloun, M., 2006]

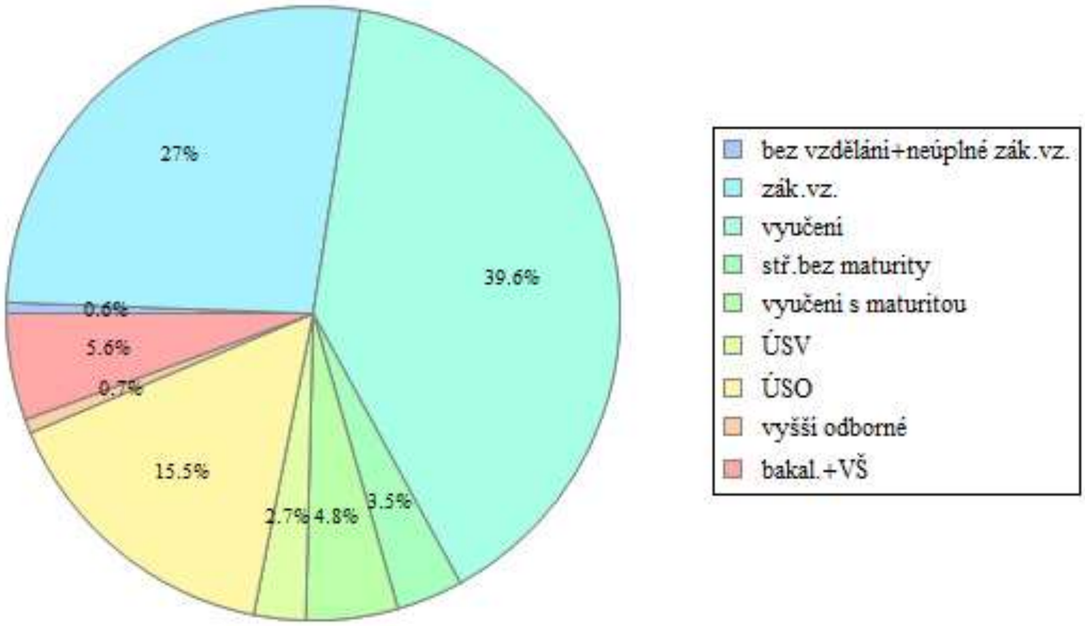
Kruhový graf většinou zobrazuje podíl jednotlivých částí na celku. [dle trilobyte]

Tabulka 4 - Struktura nezaměstnaných podle vzdělání v ČR na konci IV. čtvrtletí 2011

bez vzdělání + neúplné zák. vz.	zák. vz.	vyučení	stř. bez maturity	vyučení s maturitou	ÚSV	ÚSO	vyšší odborné	bakal. + VŠ
3001	136615	201212	17787	24676	13834	79054	3746	28526

[dle Statistická ročenka trhu práce v ČR v roce 2011]

Graf 10 - Kruhový graf struktury nezaměstnaných podle vzdělání v ČR na konci IV. čtvrtletí 2011



4.2 GRAFY VÍCEROZMĚRNÝCH DAT

Zobrazení vícerozměrných dat je možné nejen v dvourozměrném souřadnicovém systému ale i v trojrozměrném souřadnicovém systému což umožňuje
rozeznávání vektorů xi nebo jejich složek, které vypadají jako vybočující
určení struktur vyskytujících se v datech jako např. shluky, které poukazují na nehomogenitu výběru

Většinu zobrazení vícerozměrných dat je možné rozdělit do dvou základních skupin, a sice na zobecněné rozptylové diagramy a symbolové grafy.
[Meloun, M., 1998]

Ve své práci bych se zmínila pouze o grafech symbolových.

4.2.1 Symbolové grafy

Umožňují rychlé posuzování podobnosti jednotlivých objektů, jejichž proměnné jsou šifrovány do geometrických tvarů. Každý tvar tzv. symbol představuje jiný objekt. Vlastnosti těchto objektů jsou posuzovány na základě vizuálních rozdílů mezi těmito objekty.

Blíže se zaměříme na symbolové grafy - profily, polygony, tváře a stromy.
[Meloun, M.,2006]

4.2.1.1 Profily

Profily jsou dvourozměrné zobrazení m-rozměrných objektů, z nichž je každý charakterizován m proměnnými, které se zobrazují vertikálními úsečkami. Velikost těchto úseček je úměrná hodnotě odpovídající proměnné x_{ij} , j= 1,…m. Pokud spojíme koncové body úseček, vznikne profil.
[Meloun, M., 2006]

4.2.1.2 Polygony

Jako polygony můžeme nazvat profily v polárních souřadnicích. Každá proměnná objektu x^T_i, i= 1, …, n, odpovídá délce paprsku, který vychází ze společného středu.
Mezi polygony patří graf slunečních paprsků a hvězdnicový graf.
[Meloun, M., 2006]

Příkladem využití těchto typů grafů je například porovnávání kvality určitých typů výrobku např. cena, paměť, frekvence atp. kdy tyto charakteristiky jsou uvedeny na „paprsku“ o příslušné délce. Každému výrobku přísluší pro rozlišení jiná barva.

4.2.1.2.1 Graf slunečních paprsků

Graf slunečních paprsků se skládá z paprsků, které začínají ve stejném bodě, a úseček, které tyto paprsky spojují. Tvoří tak polygon. Jako společný střed paprsků se většinou volí počátek souřadnic. Každá proměnná x_{ij} objektu x^T_i odpovídá délce paprsku, které jsou ve stejných vzdálenostech rozmístěny na kružnici.

Porovnávání polygonů slouží k posouzení podobnosti objektů. Pokud je ale velké množství proměnných, stává se graf nepřehledným.

[Meloun, M., 2006]

4.2.1.2.2 Hvězdicový graf

Tento graf je podobný grafu předcházejícímu. Tvoří jej paprsky, které charakterizují relativní hodnoty proměnných u jednotlivých objektů. Paprsky se pro každý objekt spojují ve společném bodě. Paprsky stejného směru se u rozdílných objektů liší délkou.

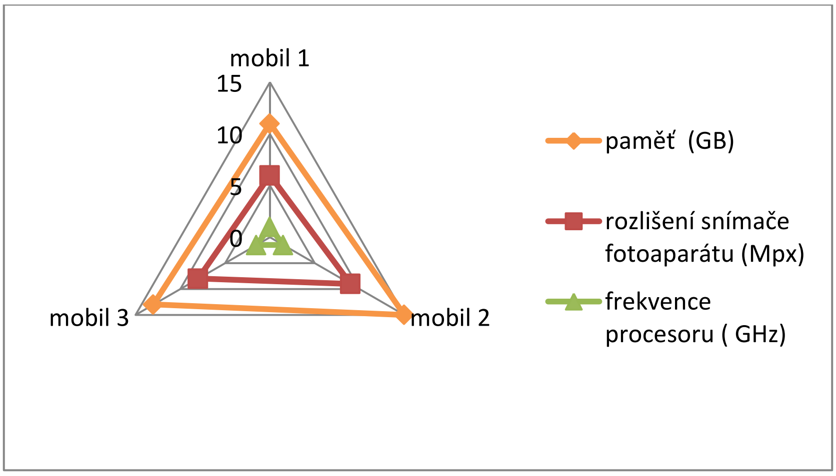
[Meloun, M., 2006]

Následující graf znázorňuje porovnání paměti, rozlišení snímače fotoaparátu a frekvence u tří rozdílných mobilních telefonů.

Tabulka 5 - Parametry mobilních telefonů

	Mobilní telefon 1	Mobilní telefon 2	Mobilní telefon 3
Paměť (GB)	11	15	13
Rozlišení snímače fotoaparátu (Mpx)	6	9	8
Frekvence **procesoru (GHz)**	1	1,5	1,5

Graf 11 - Graf slunečních paprsků



4.2.1.3 Stromy (dendogramy)

Jsou vhodné, když je počet proměnných m objektu x^T_i velký. Jednotlivé složky x_j představují délku „větví“ stromu. Struktura stromu vzniká předběžnou hierarchizací shlukování proměnných (tzv. shlukovou analýzou).

[Meloun, M., 1998]

Stromy se využívají například při Huffmanově kódování. Výsledným grafem je právě strom, který pomáhá kódovat a zpětně dekódovat binární znaky.

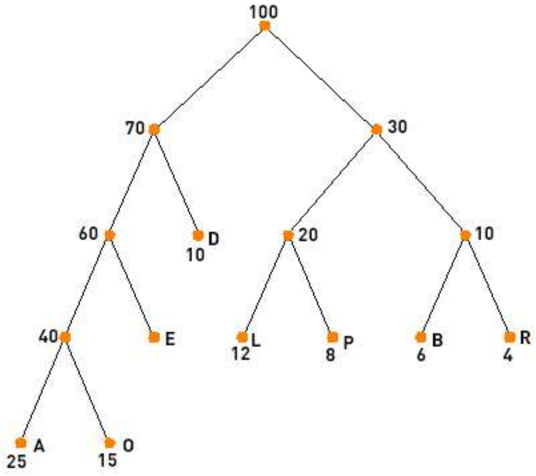
V Tabulce 6 jsou uvedena písmena s příslušnými četnostmi. Zakódujme slovo BEDLA pomocí Huffmanova algoritmu do binárních znaků.

Tabulka 6 - Písmena a jejich četnosti

A	B	D	E	L	O	P	R
25	6	10	20	12	15	8	4

Graf 12. představuje strom pro zakódování slova BEDLA do binárních znaků. Od vrcholu se snažíme dostat nejkratší možnou cestou k jednotlivým písmenům, přičemž pohyb doprava znamená hodnotu 1, pohyb doleva je hodnota 0. Tímto způsobem zakódujeme do binárních znaků všechna písmena a poté sestavíme požadované slovo.

Graf 12 - Strom Huffmanova kódu



Tabulka 7 - Zakódovaná písmena v binárních znacích

A	B	D	E	L	O	P	R
0000	110	01	001	100	0001	101	111

BEDLA – 110 001 01 100 0000

[dle Tomáš Kaiser]

4.2.1.4 Tváře

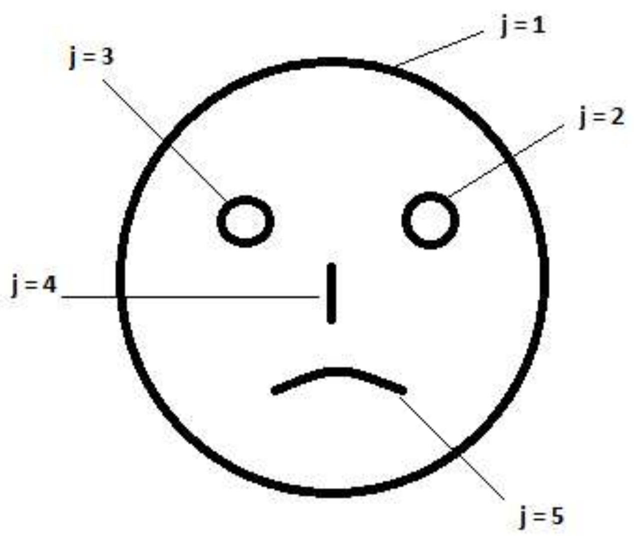
Charakterizují každou proměnnou x_{ij} objektu x^T_i jiným znakem. Jako znak můžeme brát tvar nosu, velikost nosu, velikost očí, tvar úst atp. Tvar tváře potom závisí na pořadí jednotlivých proměnných.

[Meloun, M., 1998]

Tyto grafy pomáhají rozpoznat podobnosti nebo naopak odlišnosti dat.

Na obrázku je znázorněna tvář pro 5 proměnných. Jednotlivé proměnné j potom mohou znázorňovat například charakteristiky zemí, kdy j=1 (služby), j=2 (doprava), j=3 (nezaměstnanost), j=4 (cestovní ruch), j=5 (zemědělství).

Graf 13 - Tvář



[dle Meloun, M.]

5 VYUŽITÍ GRAFŮ

Grafy jsem si pro vysvětlení jejich využití rozdělila na všeobecné a profesionální.

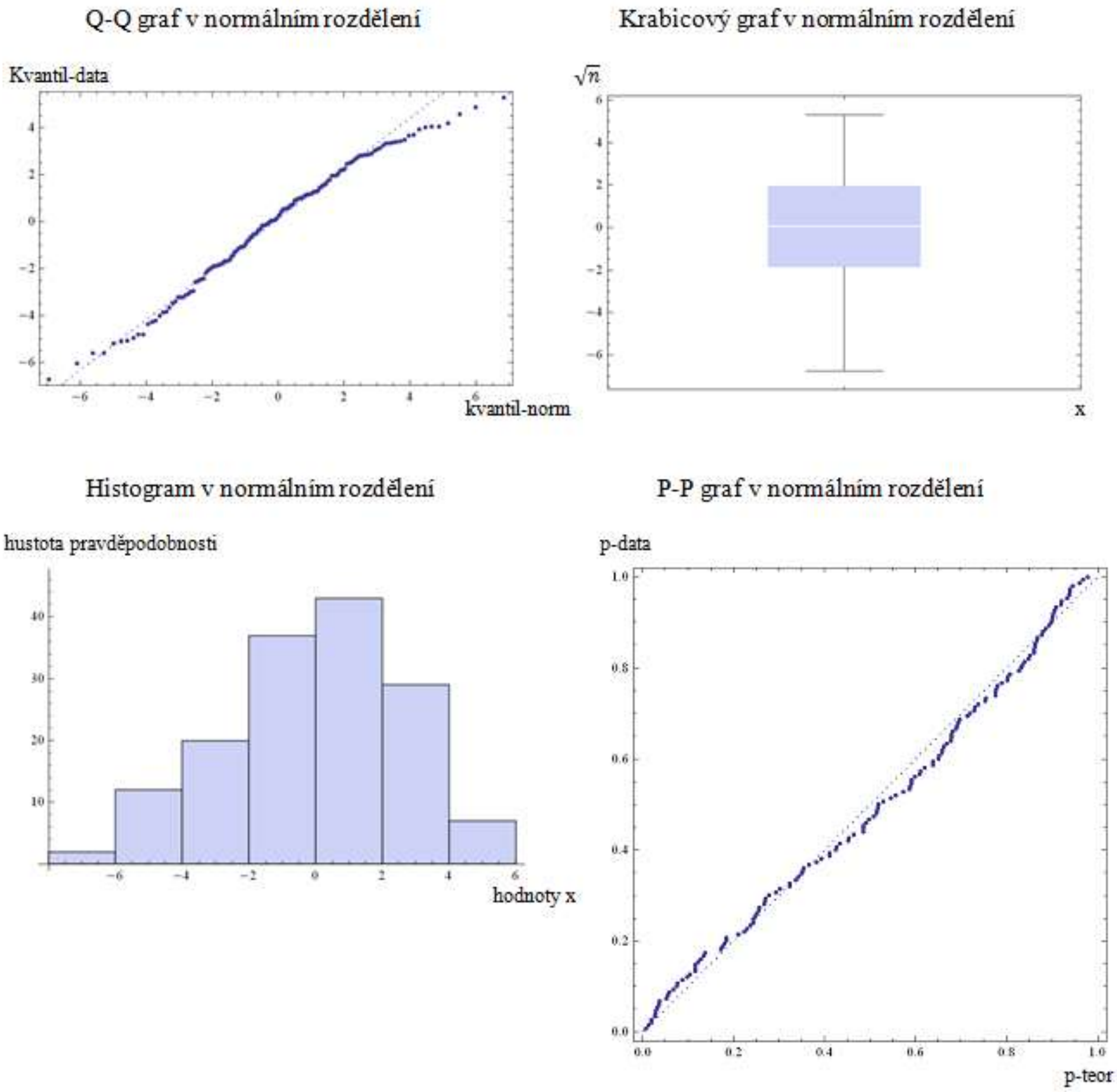
Všeobecné grafy jsou jednodušší a častěji využívané. Ve své práci jsem zmiňovala kruhový graf. Ten je jedním z nejvyužívanějších grafů vůbec. Každý si jistě vybavíme oblíbené koláče sledovanosti televizních stanic. Kruhový graf velmi přehledně prezentuje zjištěná data a je celkem jednoduché ho sestrojit. Jako další grafy bych sem zařadila ty, kterými se ale ve své práci nezabývám - například spojnicový graf nebo sloupcový graf. Všeobecné grafy se hojně vyskytují v produktech Českého statistického úřadu (ČSÚ), který je největším zdrojem dat o obyvatelstvu v České republice. Dalším místem, kde se s těmito grafy často setkáváme, je tisk nebo internetové a televizní zpravodajství. Mnoho grafů se na lidi hrne ve volebním období. Dochází k porovnávání úspěšnosti stran, graficky se znázorňují i počty voličů například s porovnáním s předcházejícími lety. Další využití je při řešení různých algoritmů, kódování, což bylo uvedeno v kapitole 3 – Grafy. Jako příklad byl uváděn graf stromu. Ten je využívám dále také při sestavování rodokmenů. Pomocí grafů je zaznamenáváno například i kmitání nebo vlnění, což využívají vědy jako fyzika a chemie. Záznam pomocí grafických přístrojů se využívá v meteorologii a mnoha dalších vědních oborech.

Profesionálními grafy jsem nazvala ty méně obvyklé, sloužící pro hlubší zkoumání dat. Tedy grafy, které zkoumají normalitu dat ve statistickém výběru. Jsou to grafy, které se využívají především při hlubších analýzách a obyvatelstvu nejsou často prezentovány. Do profesionálních grafů bych zařadila Q-Q graf, histogram, P-P graf a krabicový graf.

Grafy na následujících dvou obrázcích (Obrázek 3, Obrázek 4) znázorňují data v normálním a exponenciálním rozdělení. Z grafů jsou patrné rozdíly mezi daty těchto rozdělení. Pomocí profesionálních grafů jsme schopni analyzovat chování dat jednotlivých rozdělení, vybočující body atd.

S histogramem se na rozdíl od zbylých uvedených profesionálních grafů setkáváme častěji. Využívá se např. při zpracování věkové struktury obyvatelstva. Data jsou pak prezentována v tzv. věkové pyramidě, která je příkladem zpracování histogramu pro vícerozměrná data. Věková pyramida se skládá vlastně ze dvou histogramů, které znázorňují rozdělení podle pohlaví. Počet jednotlivých sloupců odpovídá počtu rozdělení do věkových intervalů. Následující odkaz vás přesune na stránky ČSÚ, kde je pomocí věkové pyramidy zaznamenán vývoj obyvatelstva od roku 1992 do roku 2012 (http://www.czso.cz/animgraf/cz032/).

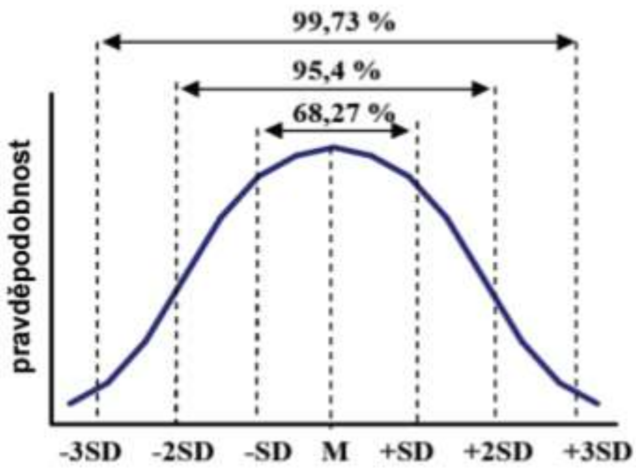
Obrázek 3 - Srovnání grafů v normálním rozdělení



Všechny grafy na Obrázku… ukazují i u jinak symetrického rozdělení lehkou asymetrii. U Q-Q grafu je patrná citlivost v oblasti chvostů, ve střední části se graf jeví podobný přímce, což se neslučuje s P-P grafem, který ukazuje na lehkou asymetrii i v oblasti střední části. Na asymetrii těchto dat ukazuje i histogram a krabicový graf, který navíc v zadaném intervalu nevykazuje žádné vybočující body (outliers).

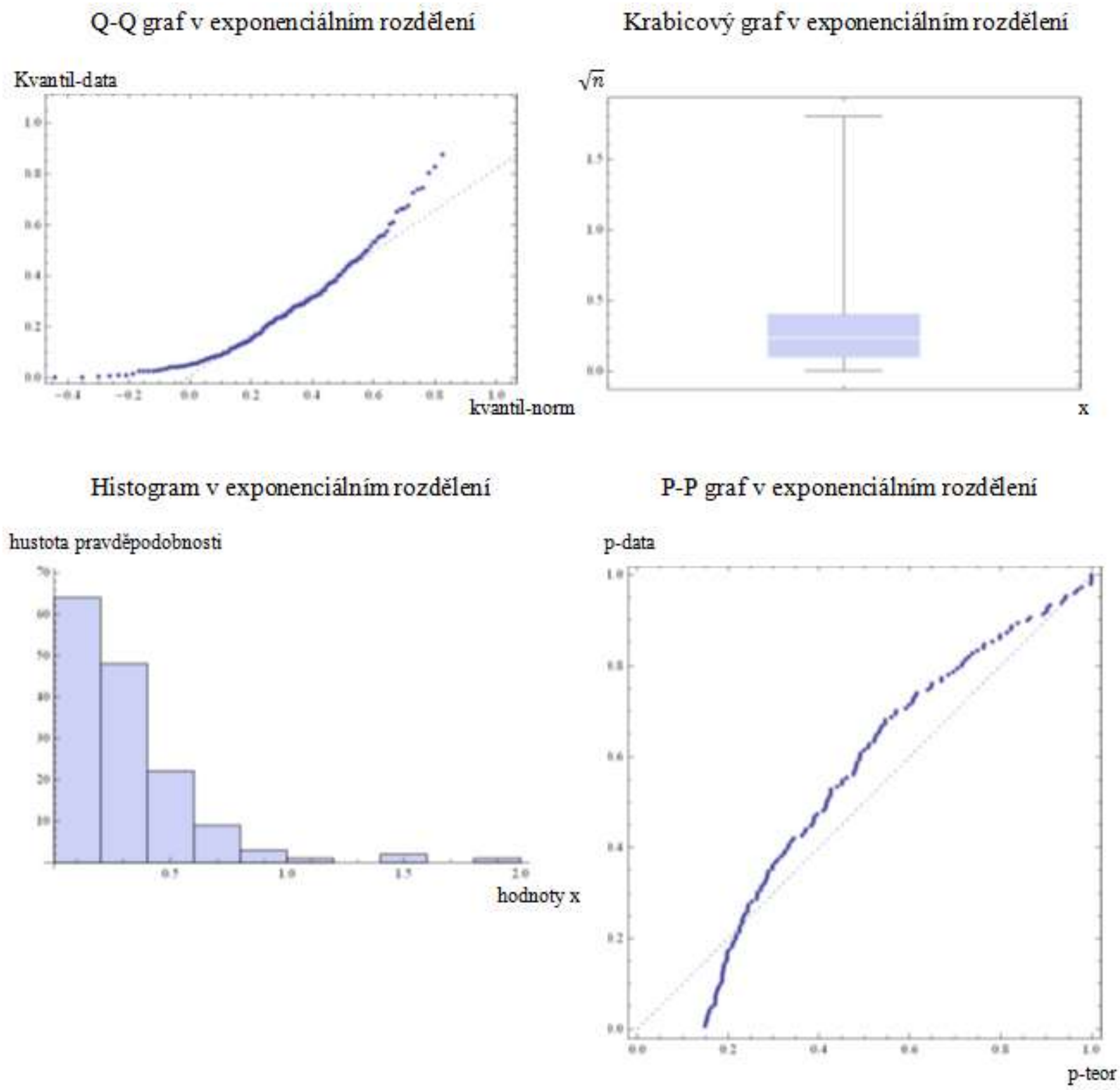
Normální rozdělení je nejčastěji používané rozdělení. Při výzkumu se setkáváme s prvky, které jsou ovlivňovány více současně působícími vlivy. Toto působení se projevuje hromaděním výsledků kolem střední hodnoty, naopak směrem od střední hodnoty je výskyt výsledků stále řidší. Toto rozdělení je také nazýváno Gaussovou křivkou. Jedná se o křivku, kterou lze sestrojit pomocí aritmetického průměru M a směrodatné odchylky SD. Pro normální rozdělení dat platí, že v intervalech (M-SD,M+SD) se vyskytuje přibližně 68,27 % hodnot, v (M-2SD, M+2SD) se vyskytuje 95,4 % hodnot a v intervalu (M-3SD, M+3SD) 99,73 % hodnot. [Základy statistiky]

Obrázek 4 - Gaussova křivka



[dle Základy statistiky]

Obrázek 5 - Srovnání grafů v exponenciálním rozdělení



Všechny typy grafů ukazují na výrazné zešikmení. Z Q-Q a P-P grafů je zřetelné, jak se tyto dva grafy doplňují. Data exponenciálního rozdělení Q-Q grafu jsou citlivá v oblastech chvostů, kdyžto data P-P grafu jsou citlivá v oblasti středu. Všechny tyto grafy celkem jasně ukazují na zobrazení, ze kterého pocházejí. Jsou velmi asymetrické. U krabicového grafu nezaznamenáváme žádné odlehlé body.

Exponenciální rozdělení je z jedné strany zdola ohraničené. Popisuje řady reálných dějů. Patří mezi rozdělení asymetrická a většinou vykazuje šikmost.

[Meloun, M., 2006]

6 ZÁVĚR

Hlavním cílem mé práce bylo seznámit čtenáře se statistickými grafy. Snažila jsem se práci vytvořit přehledně a chronologicky tak, aby na sebe jednotlivé kapitoly logicky navazovaly. Od definování statistických pojmů, které jsem využívala v dalších kapitolách, přes konkrétní definované grafy, jejich znázornění a příklady po využití těchto grafů. Nejvíce jsem pracovala se čtyřmi grafy – histogramem, Q-Q grafem, P-P grafem a krabicovým grafem. Tyto grafy jsem uváděla v normálním, rovnoměrném a exponenciálním rozdělení. V poslední kapitole, týkající se využití grafů, jsem se u těchto čtyř zmíněných grafů zaměřila na jejich porovnávání mezi sebou v normálním a exponenciálním rozdělení. Grafy byly rozděleny do dvou tříd na všeobecné a profesionální. Všeobecné grafy jsou mezi lidmi běžně využívané, proto se s nimi často setkáváme. Grafy profesionální jsou určeny pro hlubší zkoumání dat a využívají je spíše statistici.

Grafy jsem vytvářela v prostředí programu Mathematica verze 8.0 a Microsoft Excel 2010. Excel je běžně dostupným a velmi rozšířeným programem pro tvorbu grafů. Program Mathematica je známý méně. Účelem práce ale nebylo seznámit čtenáře s těmito programy, ale s již zmiňovanými statistickými grafy.

Doufám, že jsem cíl své práce splnila a čtenářům bude problematika statistických grafů jasnější než před jejím otevřením.

RESUMÉ

This bachelor’s work deals with statistical graphics and their application. The targets of the work are set in its introductory part.

“Statistical Terms” were defined in Chapter 2.

Chapter 3 describes individual types of graphs stratified into two basic groups concerning graphs with one-dimensional and multi-dimensional data.

Individual types of graphs are defined here, and demonstrated with given examples. Concluding part of this bachelor’s work shows practical application of graphs. The purpose of the work is to acquaint the reader with the possibilities and applications of statistical graphs and to achieve the final objective.

SEZNAM POUŽITÉ LITERATURY A PRAMENŮ

ČSÚ[online]. ©2014[cit. 15. 3. 2014]. Dostupné z: <http://www.czso.cz/csu/2013edicniplan.nsf/kapitola/1413-13-r_2013-10>

ČSÚ[online]. ©2014[cit. 15. 3. 2014]. Dostupné z: <http://www.czso.cz/animgraf/cz032/>

HENDL, Jan. *Přehled statistických metod.* Vydání 3. Praha: Portál, 2009. ISBN 978-80-7367-482-3

Huffmanovo kódování, Interview s Doc. RNDr. Tomášem Kaiserem, Ph.D., kantor FAV ZČU v Plzni. Plzeň 6. 5. 2013.

iba.muni. [online]. ©2014[cit. 4. 12. 2013]. Dostupné z: WWW: http://www.iba.muni.cz/esf/res/file/bimat-prednasky/biostatistika-pro-matematickou-biologii/BpMB-03.pdf

MELOUN, Milan a MILITKÝ, Jiří. *Statistické zpracování experimentálních dat*. Vydání 2. Praha: East Publishing a.s, 1998. ISBN 80-7219-003-2

MELOUN, Milan a MILITKÝ, Jiří. *Kompendium statistického zpracování dat*. Vydání 2. Praha: Academia, 2006. ISBN 80-200-1396-2

MELOUN, Milan a MILITKÝ, Jiří. *Interaktivní statistická analýza dat*. Vydání 3. Praha: Karolinum, 2012. ISBN 978-80-246-2173-9

Management Mania. [online]. ©2014[cit. 23. 7. 2013]. Dostupné z WWW: <http://www.metodyrizeni.cz/index.php/zakladni-pojmy/76-data>.

Managemen marketingu – učivo. [online]. ©2014[cit. 23. 7. 2013]. Dostupné z WWW: <http://management-marketingu.blogspot.cz/2010/09/6-primarni-sekundarni-zdroje-dat.html >

Kohout, V. Základní statistické pojmy. [online]. ©2014[cit. 11. 3. 2014]. Dostupné z: WWW: < http://www.kmt.zcu.cz/person/Kohout/info_soubory/letnisem/zs/stat9.pdf>

PROCHÁZKOVÁ, Eva a ŠEBESTOVÁ, Lucie. *Statistická ročenka trhu práce v ČR v roce 2011*. Vydání 1. Plzeň: TISKÁRNA BÍLÝ SLON s. r. o., 2012. ISBN 978-80-7421-043-3

ŠPÍRKOVÁ, Ivana. *Statistické grafy.* Plzeň, 2011. Bakalářská práce. Západočeská univerzita v Plzni, Pedagogická fakulta. [online].[cit. 29. 1. 2014]. Dostupné z WWW: <https://portal.zcu.cz/wps/portal/prohlizeni >

Štěpáníková a Čermák. Vztah mezi kvalitativním a kvantitativním výzkumem. [online]. ©2014[cit. 20. 9. 2013]. Dostupné z WWW: <abudehur.wz.cz/psycho/chovancik/metody/mvp.doc>.

Trilobyte. [online]. ©2014[cit. 23. 3. 2014]. Dostupné z WWW: <http://www.trilobyte.cz/downloadfree/qcemanual/plotting.pdf>

Základy statistiky (Skripta 2). [online]. ©2014[cit. 11. 11. 2013]. Dostupné z WWW: < files.cfkr.eu/200000080-0f29110223/ZAKLADYstatistikySKRIPTA2.pdf>.

SEZNAM GRAFŮ

ZDROJE GRAFŮ

Graf 1-10,12

zpracování v programu Mathematica 8.0

Graf 11

zpracování v programu Microsoft Excel 2010

Graf 13

MELOUN, Milan a MILITKÝ, Jiří. *Kompendium statistického zpracování dat*. Vydání 2. Praha: Academia, 2006. ISBN 80-200-1396-2

SEZNAM OBRÁZKŮ

ZDROJE OBRÁZKŮ

Obrázek 1

MELOUN, Milan a MILITKÝ, Jiří. *Interaktivní statistická analýza dat*. Vydání 3. Praha: Karolinum, 2012. ISBN 978-80-246-2173-9

Obrázek 2

vlastní zpracování

Obrázek 4

Základy statistiky (Skripta 2). [online]. ©2014[cit. 11. 11. 2013]. Dostupné z WWW: < files.cfkr.eu/200000080-0f29110223/ZAKLADYstatistikySKRIPTA2.pdf>.

Obrázek 3,5

zpracování v programu Mathematica 8.0

SEZNAM TABULEK

ZDROJE TABULEK

Tabulka 1

Kohout, V. Základní statistické pojmy. [online]. ©2014[cit. 11. 3. 2014]. Dostupné z: WWW: < http://www.kmt.zcu.cz/person/Kohout/info_soubory/letnisem/zs/stat9.pdf>

Tabulka 2,3,5

Vlastní zpracování

Tabulka 4

PROCHÁZKOVÁ, Eva a ŠEBESTOVÁ, Lucie. *Statistická ročenka trhu práce v ČR v roce 2011*. Vydání 1. Plzeň: TISKÁRNA BÍLÝ SLON s. r. o., 2012. ISBN 978-80-7421-043-3

Tabulka 6,7

Huffmanovo kódování, Interview s Doc. RNDr. Tomášem Kaiserem, Ph.D., kantor FAV ZČU v Plzni. Plzeň 6. 5. 2013.