

Vysoká škola ekonomická v Praze
Fakulta managementu v Jindřichově Hradci

Diplomová práce

Martin Baumann
2017



Vysoká škola ekonomická v Praze

Fakulta managementu v Jindřichově Hradci

Katedra exaktních metod

Statistika a její využití (zneužití) při tvorbě informací

Autor diplomové práce:	Bc. Martin Baumann
Vedoucí diplomové práce:	doc. RNDr. Jitka Bartošová, Ph.D.
Rok obhajoby:	2017

Čestné prohlášení

*Prohlašuji, že diplomovou práci na téma
„Statistika a její využití (zneužití) při tvorbě informací“
jsem vypracoval samostatně a veškerou použitou literaturu a další prameny
jsem řádně označil a uvedl v přiloženém seznamu.*

V Jindřichově Hradci dne 30. června 2017

podpis



ZADÁNÍ DIPLOMOVÉ PRÁCE

Autor práce: Bc. Martin Baumann
Studijní program: Ekonomika a management
Obor: Management

Vedoucí práce: doc. RNDr. Jitka Bartošová, Ph.D.

Název práce: **Statistika a její využití (zneužití) při tvorbě informací**

Jazyková varianta: Čeština

Rámcový obsah:

1. Cílem práce je popis využití statistických ukazatelů a metod k tvorbě informací a ovlivňování veřejného mínění.

Rozsah práce: 60

Literatura:

1. Seznam literatury bude součástí diplomové práce

Datum zadání: leden 2016

Datum odevzdání: duben 2017

Bc. Martin Baumann
Řešitel

doc. RNDr. Jitka Bartošová,
Ph.D.
Vedoucí práce

Ing. Vladimír Příbyl, Ph.D.
Vedoucí katedry

doc. Ing. Vladislav Bína, Ph.D.
Děkan FMJH VŠE

Název diplomové práce:

Statistika a její využití (zneužití) při tvorbě informací

Abstrakt:

Veřejné mínění je utvářeno v několika vzájemně provázaných fázích. Na počátku však vždy stojí specifická informace nebo událost, které je daný člen společnosti nebo část populace vystavena. Neustále se zvyšující možnosti získávání, zpracování a analýzování dat dávají prostor k vytváření informací o různých oblastech lidské činnosti. Tyto informace hrají v procesu tvorby veřejného mínění významnou roli. K účelům analýzy dat jsou nejčastěji využívány různé prvky statistiky. Diplomová práce se zabývá možnostmi využití statistiky při tvorbě a předávání informací. Teoretická část práce se zabývá poznatky z oblasti veřejného mínění a statistiky. V rámci praktické části práce jsou představeny nejčastější postupy, které jsou využívány k záměrné či nevědomé manipulaci s veřejným míněním. Ústředním prvkem praktické části práce je sestavení dotazníkového šetření a analýza z něho pocházejících dat. Kardinální výzkumnou hypotézou diplomové práce je zjištění, zda ovlivňuje znalost statistických pojmů schopnost respondenta interpretovat informační sdělení.

Klíčová slova:

Veřejné mínění, statistika, zneužití statistiky, média.

Poděkování:

Rád bych poděkoval vedoucí diplomové práce paní doc. RNDr. Jitce Bartošové, Ph.D. za cenné rady a připomínky, které jsem v průběhu zpracování diplomové práce dostával. Poděkování patří také všem respondentům, kteří se zúčastnili dotazníkového šetření.

Obsah

Úvod	15
1 Veřejné mínění	17
1.1 Tvorba veřejného mínění	18
1.2 Průzkumy veřejného mínění	19
1.2.1 Metody průzkumů	21
2 Statistika	25
2.1 Statistická analýza dat	26
2.1.1 Analýza jednotlivých proměnných	26
2.1.2 Testování hypotéz	33
3 Metodologie diplomové práce	37
3.1 Dotazníkové šetření	37
4 Jak „lhát“ se statistikou	40
4.1 Problém nereprezentativního vzorku	41
4.2 Zavádějící střední hodnoty	43
4.3 Změť grafů	46
4.4 Klam narativity	50
5 Dotazníkové šetření a analýza dat	52
5.1 Výsledky dotazníkového šetření	52
5.1.1 Charakteristika respondentů	53
5.1.2 Analýza odpovědí na základní statistické pojmy	57
5.1.3 Analýza schopnosti interpretace informačních sdělení	60
5.2 Analýza závislostí	63
5.2.1 Závislost znalosti statistických pojmů na daných proměnných	63
5.2.2 Závislost schopnosti interpretovat zprávy na znalosti statistických pojmů	67
5.3 Shrnutí výsledků dotazníkového šetření	68
Závěr	71
Seznam literatury	74
Přílohy	78
Příloha č. 1: Výsledky statistických testů	78
Příloha č. 2: Dotazníkové šetření	87

Seznam obrázků

Obrázek 1 – Faktory ovlivňující veřejné mínění	19
Obrázek 2 – Fáze průzkumu veřejného mínění.....	21
Obrázek 3 – Vývoj znalostí.....	25
Obrázek 4 – Normální rozdělení s průměrem 10	31
Obrázek 5 – Normální rozdělení s průměrem 10	31
Obrázek 6 – Vztah chyby I. a II. druhu	35
Obrázek 7 – Typologie dotazníků	38
Obrázek 8 – Zdroje chyb a možnosti klamání	40
Obrázek 9 – Průměrný počet zbraní na 100 obyvatel	45
Obrázek 10 – Vývoj amerických domácností vlastníků zbraně	45
Obrázek 11 – Graf prezentovaný na zasedání tripartity dne 18. května 2015.....	47
Obrázek 12 – Příklad zavádějícího formátu koláčového grafu	49
Obrázek 13 – Korelace neimplikující kauzalitu.....	50

Seznam tabulek

Tabulka 1 – Charakteristiky kvantitativního, kvalitativního a smíšeného výzkumu	24
Tabulka 2 – Schéma rozdělení četností.....	27
Tabulka 3 – Chyby I. a II. druhu	35
Tabulka 4 – Pravděpodobnosti chyb I. a II. druhu	35
Tabulka 5 – P-hodnota a hypotézy	36
Tabulka 6 – Výsledky průzkumu prezidentských voleb USA v roce 1936	41
Tabulka 7 – Výsledky prezidentských voleb USA v roce 1936	42
Tabulka 8 – Výsledky průzkumu prezidentských voleb USA v roce 2016	43
Tabulka 9 – Zdroje pro výrobu energie v USA v roce 2014	49
Tabulka 10 – Pohlaví respondentů	53
Tabulka 11 – Věkové složení respondentů.....	54
Tabulka 12 – Nejvyšší dosažené vzdělání.....	54
Tabulka 13 – Status respondentů.....	55
Tabulka 14 – Respondenti podle oboru.....	56
Tabulka 15 – Bydliště respondentů.....	57
Tabulka 16 – „Kolik je průměr?“	58
Tabulka 17 – „Kolik je modus?“	58
Tabulka 18 – „Kolik je medián?“	59
Tabulka 19 - „Co je to rozptyl?“	59
Tabulka 20 – Znalost základních statistických pojmů.....	60
Tabulka 21 - „Odráží novinový titulek skutečnost?“	61
Tabulka 22 – „Je na základě provedeného průzkumu možno spolehlivě určit míru životní spokojenosti občanů ČR?“	61
Tabulka 23 – Schopnost interpretovat koláčový graf.....	62
Tabulka 24 – Schopnost interpretovat sloupcový graf	62
Tabulka 25 – Schopnost interpretace informačních sdělení.....	63
Tabulka 26 – Znalost statistických pojmů	64
Tabulka 27 – Znalost statistických pojmů po úpravě	65
Tabulka 28 – Četnost věkových kategorií po úpravě.....	65
Tabulka 29 – Bydliště respondentů dle historických území ČR.....	65
Tabulka 30 – Stupeň dosaženého vzdělání po úpravě	66
Tabulka 31 - Četnost statusu po úpravě	67
Tabulka 32 – Znalost statistických pojmů po úpravě	68
Tabulka 33 – Schopnost interpretace informací po úpravě.....	68
Tabulka 34 - Kontingenční tabulka proměnných znalost x pohlaví.....	78
Tabulka 35 - Chi-kvadrát test nezávislosti (znalost x pohlaví)	78

Tabulka 36 – Kontingenční tabulka proměnných znalost x věková kategorie	79
Tabulka 37 - Kontingenční tabulka proměnných znalost x věková kategorie po úpravě	80
Tabulka 38 – Chí-kvadrát test nezávislosti (znalost x věk)	80
Tabulka 39 - Kontingenční tabulka proměnných znalost x bydliště.....	81
Tabulka 40 - Kontingenční tabulka proměnných znalost x bydliště po úpravě	82
Tabulka 41 - Chí-kvadrát test nezávislosti (znalost x bydliště)	82
Tabulka 42 - Kontingenční tabulka proměnných znalost x vzdělání.....	83
Tabulka 43 - Kontingenční tabulka proměnných znalost x vzdělání po úpravě.....	83
Tabulka 44 - Chí-kvadrát test nezávislosti (znalost x vzdělání)	83
Tabulka 45 - Kontingenční tabulka proměnných znalost x status po částečné úpravě	84
Tabulka 46 - Kontingenční tabulka proměnných znalost x status po konečné úpravě	84
Tabulka 47 - Chí-kvadrát test nezávislosti (znalost x status)	84
Tabulka 48 - Kontingenční tabulka proměnných schopnost interpretace x znalost pojmů	85
Tabulka 49 - Kontingenční tabulka proměnných schopnost interpretace x znalost pojmů po úpravě	85
Tabulka 50 - Chí-kvadrát test nezávislosti (schopnost interpretace x znalost pojmů)	86

Seznam grafů

Graf 1 – Vývoj minimální mzdy v ČR od roku 2007	48
Graf 2 – Struktura respondentů podle pohlaví.....	53
Graf 3 – Paterův graf rozdělení respondentů podle bydliště	56

Úvod

Téma diplomové práce, které budu v rámci magisterského studia zpracovávat, nese název „*Statistika a její využití (zneužití) při tvorbě informací*“. Motivací pro výběr tohoto tématu byl především můj osobní zájem o statistiku jako celek a zejména problematiku předávání informací, které ze statistiky vyplývají, veřejnosti.

Téměř každý člověk se denně, ať cíleně, či náhodou, dostává do konfrontace s informacemi, které číhají prakticky na každém rohu. Při pouhé cestě do práce jsme vystavováni informačním sdělením z tisku, reklamních billboardů nebo rádia. Pomocí průměrů, mediánů, grafů, procent nebo jiných statistických pojmů se nás tyto prostředky snaží přesvědčit o nějaké skutečnosti. Tento příliv na nás mnohdy působí spíše formou, kterou je nám předkládán než samotným obsahem. Konzument informací si v té záplavě ani nestihne uvědomit, z jakých faktů dané sdělení vychází nebo jakým způsobem se k nim její původce dopracoval. Velkým problémem je situace, kdy autor zprávy účelně předkládá grafy, diagramy nebo tabulky, které jsou zavádějící nebo neúplné. Takový postup pak v recipientovi informací vzbuzuje mylné závěry, které nemusí reflektovat pravou realitu. Tímto způsobem pak dochází k účelnému ovlivňování veřejného mínění, aniž by si to většina osob uvědomovala. Jedním z úkolů diplomové práce proto bude informovat čtenáře o možných způsobech manipulace s veřejným míněním a v budoucnu je donutit k větší ostražitosti při konzumaci informačního obsahu jakékoliv zprávy nebo sdělení.

Struktura diplomové práce „*Statistika a její využití (zneužití) při tvorbě informací*“ bude, jako většina prací tohoto typu, obsahovat dva hlavní celky, teoretickou a praktickou část.

Teoretická část bude rozdělena do dvou kapitol. V té první bude pozornost věnována pojmu veřejné mínění, kdy bude vysvětlen jeho význam, historické kořeny a způsoby průzkumu veřejného mínění. V druhé části se budu zabývat již samotnou statistikou. Stejně jako v části první zde bude nastíněn historický vývoj statistiky a představeny a vysvětleny základní statistické pojmy a ukazatele, které jsou důležité pro správnou interpretaci výsledků dotazníkového šetření.

Praktická část bude začínat nastíněním metodiky, která bude následně využita pro její zpracování. Zde bude popsána zejména vybraná metoda sběru dat.

První část v praktickém oddílu práce se bude zabývat praktickými příklady účelové prezentace informací v médiích a jejich analýze. Budu se snažit najít odpověď na otázku „*Jaké způsoby prezentace dat jsou zneužívány k ovlivňování veřejného mínění a v jaké formě?*“ Budou zde uvedeny některé příklady, při kterých média účelně prezentují informace.

Hlavní částí diplomové práce bude vyhodnocení dat získaných z dotazníkového šetření a nalezení odpovědí na stanovené výzkumné otázky.

Druhá výzkumná otázka, na kterou budu hledat odpověď pomocí dat získaných z dotazníku, zní: *„Na jakých faktorech závisí znalost statistických pojmů, které jsou běžně používány v médiích?“* Výsledek této otázky vyplyne z úspěšnosti, s jakou budou respondenti odpovídat na kladené dotazy týkající se statistických pojmů jako například průměr, medián, modus nebo rozptyl. Tato otázka vychází z předpokladu, že aby mohl člověk správně interpretovat předkládané informace, musí jim v první řadě rozumět.

Třetí výzkumná otázka zní: *„Závisí schopnost správné interpretace informací na znalostech statistických pojmů?“* K nalezení odpovědi na tuto otázku budou použita data také z další sekce dotazníku, ve které budou respondentům úmyslně předkládány účelně konstruované grafy a sledována úspěšnost, s jakou na ně budou odpovídat. Cílem této otázky bude zjistit, jak neznalost statistických pojmů ovlivní správnou interpretaci informací. Tato otázka je klíčová, protože pokud člověk nedokáže správně interpretovat informace, se kterými se dostává denně do styku, lze jeho mínění snáze ovlivnit.

1 Veřejné mínění

S pojmem veřejné mínění se setkáváme velmi často a většina osob mu rozumí nebo má alespoň představu o tom, co tento pojem znamená. Pouze málo osob by však tento pojem dokázalo definovat, a když už ano, jednotlivé definice by se značně lišily. S podobným problémem se potýkají i samotní autoři a specialisté z oblasti průzkumů veřejného mínění. Již při dataci konkrétního období, ve kterém se začal tento pojem objevovat poprvé, se jednotliví teoretici značně rozcházejí (Kohout, 1999).

Někteří autoři se domnívají, že tento pojem byl znám již ve starověkém Řecku a dokonce se ho někteří významní političtí řečníci té doby snažili ovlivňovat. Obdobná situace panovala i ve starověkém Římě, ze kterého se dochovala věta „*Hlas lidu, hlas boží*“, kde právě onen hlas lidu označuje veřejné mínění (Glasser a Salmon, 1995). Dokonce i v prastarém čínském písmu se objevují znaky, které mají význam slov veřejný a mínění (Svoboda, 2009). Většina autorů však považuje za vznik pojmu veřejné mínění až 18. století, kdy ve své knize s názvem „*O společenské smlouvě neboli o zásadách státního práva*“ pasuje francouzský osvícenský filozof Jean Jacques Rousseau tento výraz mezi nejdůležitější zákony demokratické společnosti (Urban et al., 2011). Ve stejném století použil tento výraz i Francouz Charles Pinot Duclos. Podle tohoto spisovatele je veřejné mínění pro společnost neméně důležité jako obživa. Lidé totiž podle něho nejsou spojováni pouze materiálovými potřebami, ale nýbrž také vzájemnými názory a stanovisky (Walton, 2009).

Dle mého názoru není přesná učebnicová definice pojmu veřejné mínění nezbytně nutná. Důležitější jsou hlavní rysy a znaky veřejného mínění, které zachycuje Kohout (1999), a to, že veřejné mínění:

- „*odráží současné názory, postoje a nálady veřejnosti;*
- *nelze ho považovat za přesné rozumové poznání;*
- *má silnou dynamiku a rychle se může měnit;*
- *vždy obsahuje prvky subjektivnosti, přibližnosti a dojmovosti;*
- *je dáno společenským zájmům, znalostí a tradic;*
- *vytváří se jen k významným podnětům (jevům, názorům, osobnostem nebo událostem) a*
- *je ovlivnitelné mnoha způsoby (projevy politiků, manipulací demagogů, masmédií nebo významných a veřejně známých osobností)* (Kohout, 1999, s. 15).“

Od doby, kdy se pojem veřejné mínění objevil, uplynulo tedy více než 250 let. Dnes má však tento výraz mnohem větší váhu než kdykoliv dříve. Může za to zejména vysoká informovanost lidí, kteří jsou díky masmédiím schopni konzumovat informace od osobností a informačních kanálů z celého světa. K rostoucímu významu veřejného

mínění přispěla také tvorba demokratických systémů, se kterými beze sporu souvisí právo svobodného vyjadřování a svobodné volby. Stačí porovnat situaci v České republice před revolucí v listopadu roku 1989, kdy svobodné vyjadřování v jiném než ideologickém duchu bylo nemilosrdně perzekuováno. Dnes však nejde pouze o činy politiků, kteří musí brát při svých rozhodnutích v úvahu názory svých nebo potenciálních voličů. Mnohem důležitější je skutečnost, že i jednotlivé firmy, které se pohybují v tržní ekonomice, musejí brát při svých činnostech ohled na mínění zákazníků (Kohout, 1999).

1.1 Tvorba veřejného mínění

Veřejné mínění není statickým jevem, ale jako spousta dalších jevů v naší společnosti prochází několika specifickými fázemi. Mišovič (2010) shrnuje celkem 4 fáze, které se vzájemně prolínají, a to:

- utváření,
- vyjádření,
- fungování,
- uplatnění.

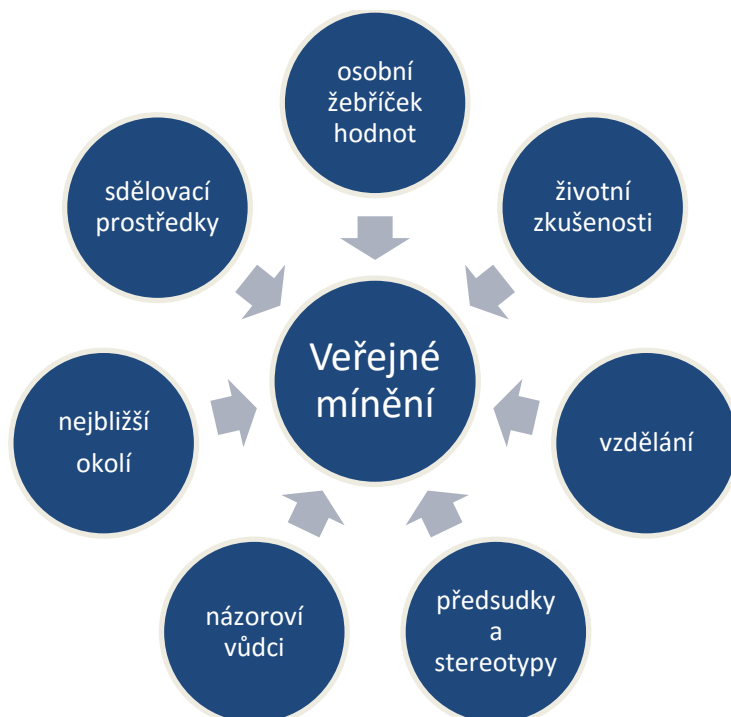
Samotný počátek veřejného mínění se vždy váže k nějaké události, která se týká velké skupiny osob. V dnešní záplavě informací může za vznikem veřejného mínění stát obyčejná zpráva v novinách. Každý jedinec má k dané věci individuální přístup, který vede ke krystalizaci veřejného mínění. V rámci té dochází k vytvoření vlastního mínění na daný problém. Vlastní názor však k vytvoření veřejného mínění nestačí. K tomu je nutné rozpoutat diskuzi, ve které se dostávají do vzájemné konfrontace individuální pohledy a dochází k formování společného úsudku dané společnosti. V tomto procesu, který je již součástí druhé fáze, se může mínění jednotlivce měnit v závislosti na jeho víře v kompetentnost zbytku společnosti nebo jeho osobnostních rysech.

Ve třetí fázi procesu tvorby veřejného mínění je samotné fungování veřejného mínění. Do této fáze jsou zahrnuty různé aktivity, kterými dává společnost najevo svá stanoviska k daným jevům. Mezi dané aktivity lze zařadit například organizování veřejných projevů a shromáždění či jiné formy prezentace skupinových názorů.

Poslední čtvrtá část souvisí zejména s direktivní a kontrolní funkcí veřejného mínění a stejně jako fáze předešlá poskytuje zpětnou vazbu na situaci v určité společnosti. Fáze uplatnění veřejného mínění zahrnuje tedy například tlak na dodržování zákonů, podávání petic nebo hlasování ve volbách. Tato fáze trvá tak dlouho, dokud je daný problém pro společnost důležitý a cítí potřebu se jím intenzivně zabývat (Mišovič, 2010).

Celý proces tvorby veřejného mínění je vždy ovlivněn specifickými faktory prostředí, ve kterém k němu dochází. Na Obrázku 1 jsou zobrazeny faktory, které se podílejí na tvorbě veřejného mínění nebo na jeho změně (Červenka a Kunštát, 2016).

Obrázek 1 – Faktory ovlivňující veřejné mínění



Zdroj: Červenka a Kunštát (2006), vlastní zpracování

Výše vlivu jednotlivých faktorů je různá a závisí zejména na typu daného problému a osobách, které se daným problémem zabývají a spoluvytvářejí veřejné mínění.

1.2 Průzkumy veřejného mínění

Ke zjištění veřejného mínění určité společnosti nebo skupiny osob se využívají průzkumy veřejného mínění. Ty se snaží zjistit názory populace nebo častěji jejího vzorku na určitý problém nebo jev.

Historie průzkumů veřejného mínění je pravděpodobně stejně stará jako pojem veřejné mínění samotný. Již ve starověkém Řecku docházelo k hlasování občanů o vyhoštění nebezpečných lidí ze společnosti. Další zmínky o tomto oboru pocházejí ze 17. století, kdy docházelo k účelnému sběru dat o obyvatelstvu Francie a jeho názoru (Novotná, 2011). Na této datové základně byla následně provedena sociálně statistická analýza. Také v Českých zemích se již v průběhu 18. století konal průzkum, který zjišťoval nálady venkovských obyvatel (Šubrt, 1998).

Další významný milník ve vývoji tohoto oboru se datuje do 20. století. Roku 1935 založil americký sociolog a statistik George Gallup Americký ústav pro výzkum veřejného mínění. Jeho přístup k výzkumu veřejného mínění z této doby odpovídal svými parametry dnešním přístupům. Hlavní dva parametry, které lze považovat za přelomové, byla technika interview a výběr reprezentativního vzorku respondentů (Urban et al., 1999).

Technika interview spočívá v důkladném proškolení tazatelů. Ti jsou před samotným šetřením poučeni o přesném postupu, kterým má být dotazování vedeno. Díky tomu jsou všichni tazatelé schopni s dotazovaným vést standardizovaný řízený rozhovor. Na jeho základě je následně možné získat zhruba stejné údaje od všech dotazovaných a dále je mezi sebou porovnávat.

Při výběru reprezentativního vzorku respondentů začal Gallup využívat výběrových statistických postupů. Tento krok mu napomohl získat ze základního statistického souboru takový výběrový soubor, který dokázal reprezentovat a podávat věrný obraz o celém základním souboru. „*Statistické výběrové postupy jsou pro něj symbolem a současně i zárukou vědeckosti a objektivity výzkumu* (Šubrt, 1998, s. 15).“

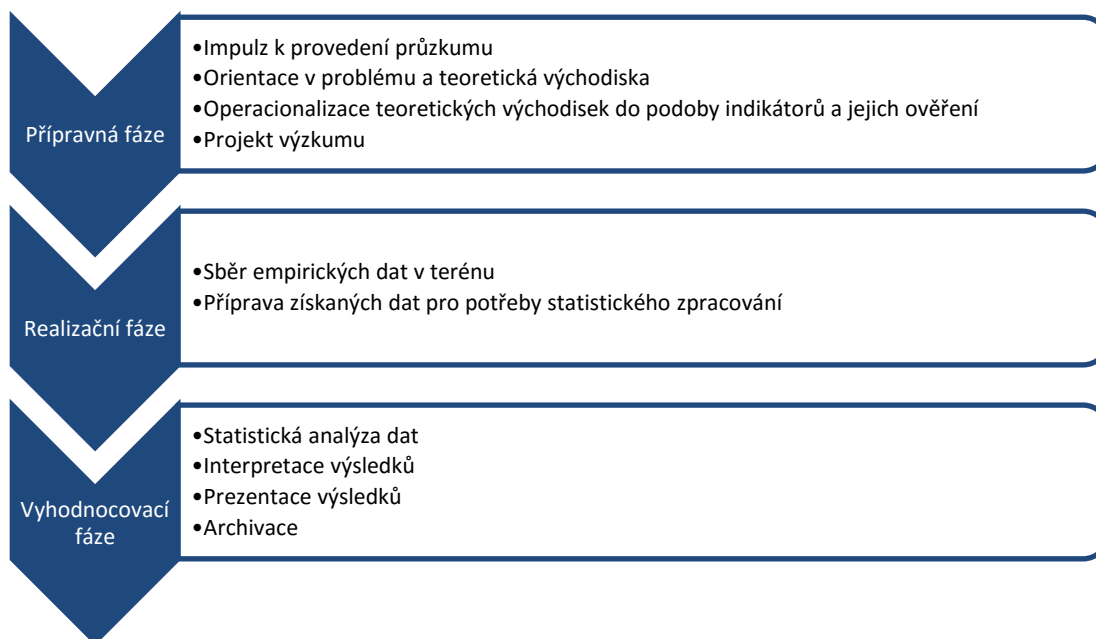
S rostoucím významem informací začalo mít mnoho subjektů zájem veřejné mínění zjišťovat a dále zkoumat. Huk (2013) uvádí 3 skupiny osob, které průzkumy provádějí. Do první skupiny řadí osoby, které průzkumy provádí primárně z důvodu touhy po určitém poznání. Do druhé skupiny patří lidé, kteří předpokládají, že se jim na základě znalostí získaných z průzkumů podaří realizovat takové kroky, které povedou k jejich kariérnímu růstu, vzestupu ve společnosti či politice. A třetí skupinu reprezentují osoby, které prahnou po podnikatelském úspěchu a pro které perfektně provedené průzkumy představují jednu z podmínek pro dosažení tohoto úspěchu.

Předmětem průzkumu se může stát prakticky jakákoliv věc, o které má výzkumník zájem získat nějaké informace. Stejná situace je i v případě volby vzorku, u kterého mají být dané věci zjišťovány. V praxi se rozlišuje následujících 5 skupin, které lze podrobit zkoumání:

- obecná populace,
- pouze určitá část obecné populace vymezená určitým společným znakem,
- skupina populace, která se vyznačuje jednotnou sociální, společenskou nebo životní situací,
- skupina populace sdílející společné stanovisko či zájem,
- cílová skupina, jež je definována zadavatelem, např. určitá část trhu (Huk, 2013).

Ačkoliv se jednotlivé průzkumy od sebe výrazně odlišují svou meritorní stránkou, ta procesní je velmi podobná. Celý proces průzkumu veřejného mínění lze proto rozdělit do několika částí. Urban et al. (2011) rozlišuje celkem 3 fáze, přípravnou, realizační a vyhodnocovací. Jednotlivé fáze společně s jejich nižšími úrovněmi jsou zobrazeny na Obrázku 2.

Obrázek 2 – Fáze průzkumu veřejného mínění



Zdroj: Urban et al. (2011), vlastní zpracování

Na začátku přípravné fáze se objevuje nějaký podnět, který vede k provedení daného průzkumu, seznámení s teoretickou problematikou a sestavení projektu výzkumu. Realizační fáze v sobě zahrnuje terénní sběr dat a jejich přípravu pro následné statistické zpracování. Ve třetí vyhodnocovací fázi dochází k analýze získaných dat a interpretaci nebo prezentaci získaných výsledků.

1.2.1 Metody průzkumů

Na začátku každého průzkumu je nutné určit statistický soubor. Jedná se o množinu jednotek nebo jedinců, u kterých budou zkoumány určité statistické znaky. Ty reprezentují vlastnosti statistických jednotek a usilují o jejich popis. Statistické znaky lze rozdělit do dvou skupin, kvalitativní a kvantitativní. Kvalitativní neboli kategoriální znaky jsou vyjádřeny slovně a ve většině případů nabývají menšího počtu obměn. Typickým příkladem tohoto typu znaku může být například pohlaví, národnost nebo

barva očí. Pokud lze tyto znaky logicky řadit dle určitého klíče jsou nazývány ordinálními. V případech, ve kterých toto řazení není možné nebo logické jde o kvalitativní nominální znaky.

Druhou skupinou jsou znaky kvantitativní neboli číselné, které v porovnání se znaky kvalitativními nabývají velkého množství možností. Kvantitativní znaky je možné rozdělit na spojité a diskrétní podle toho, zda je možné porovnávat získané hodnoty rozdílem nebo podílem. Typickým příkladem kvantitativní proměnné je výše mzdy, rok narození nebo cena produktu.

Pokud má výzkumník k dispozici kompletní množinu jednotek, je tento soubor označován jako základní. Velmi často však není možné z časových, finančních nebo technických důvodů provést průzkum u všech statistických jednotek populace. V tom případě dojde za použití specifického klíče k výběru pouze části jednotek z celkové populace. Tento reprezentativní vzorek je potom nazýván výběrovým souborem.

Průzkumy by měly usilovat o dosažení spolehlivých a validních výsledků. Zásadní podmínkou pro dosažení takových výsledků je volba vhodné metody průzkumu pro daný problém. Hlavní dva způsoby, jak dosáhnout požadovaného výsledku jsou kvalitativní a kvantitativní průzkumy. Relativně novým přístupem je pak smíšený průzkum, který vzešel z kombinace výše uvedených dvou přístupů.

Kvantitativní průzkumy

Kvantitativní průzkumy jsou založeny na paradigmatech pozitivizmu. Na základě toho se předpokládá, že existuje pouze jedna objektivní skutečnost, která je zcela nezávislá na postojích nebo názorech výzkumníka nebo výzkumného týmu. Psycholog a sociolog Fred N. Kerlinger (1972, s. 27) definuje kvantitativní vědecký výzkum jako „*systematické, kontrolované, empirické a kritické zkoumání hypotetických výroků o předpokládaných vztazích mezi přirozenými jevy*“.

Kvantitativní průzkum se tedy zabývá především frekvencemi a četnostmi daných jevů a hledáním odpovědí na otázky „*kolik?*“ nebo „*kdy?*“. Ve většině kvantitativních průzkumů je využit hypoteticko-deduktivní přístup. Na začátku deduktivního postupu dochází k formulaci obecného tvrzení, které vychází z teorie a lze jím vysvětlit vztahy prvků v reálném světě. Následně dochází k formulaci hypotéz a sběru dat, na základě kterých lze potvrdit či zamítnout dané hypotézy. Posledním krokem tohoto přístupu může být zobecnění získaných výsledků na celou populaci (Kozel et al., 2011).

Kvantitativní průzkum je prováděn systematickým postupem s předem stanovenými cíli. Huk (2013) rozeznává celkově 4 techniky využitelné k vykonávání kvantitativních průzkumů, a to:

- přímé pozorování,
- rozhovor,
- dotazník a
- analýza již vytvořených dokumentů.

V dnešní době je nejčastěji používaným nástrojem kvantitativních průzkumů dotazník, jenž bude použit také v praktické části diplomové práce.

Kvalitativní průzkumy

Kvalitativní průzkumy na rozdíl od těch kvantitativních připouští existenci více realit a klade důraz na subjektivní jednání osob. Dlouhou dobu byl tento přístup považován za pouhý doplněk ke kvantitativnímu průzkumu, dnes je však kvalitativní výzkum rovnocenným partnerem a v některých odvětvích má nezastupitelnou pozici.

Při provádění kvalitativních průzkumů dochází při sběru dat do jisté míry k improvizaci a samotný výzkumník je často považován za detektiva, protože s postupným hromaděním dat dochází k neustálému vývoji metod, otázek nebo hypotéz. Hendl (2008) ho z tohoto důvodu považuje za emergentní či pružný výzkum. Huk (2013) uvádí celkem 3 metody provádění kvalitativních průzkumů:

- zúčastněné pozorování, při kterém se výzkumník pohybuje v přirozeném prostředí sledovaného vzorku populace,
- nestandardizovaný rozhovor a
- analýza dokumentů, které se přímo týkají sledovaných osob nebo subjektů.

Smíšené průzkumy

S postupným posilováním pozice kvalitativního výzkumu se metodologové začali zabývat spojováním obou přístupů v rámci jedné studie, díky čemuž došlo ke vzniku smíšeného neboli kombinovanému průzkumu. Smíšený průzkum kombinuje výhody obou předchozích metod a umožňuje relevantnější odpovědi na stanovené výzkumné otázky. Smíšený výzkum se objevuje ve dvou základních variantách. První z variant využívá pro jednotlivé kroky výzkumu vždy pouze jednu metodu. V té druhé potom dochází ke kombinování obou metod v rámci jedné fáze.

Greene et al. (1989) na základě analýzy smíšených výzkumů stanovil celkem 5 důvodů k využívání smíšených výzkumů, a to:

- potvrzení výsledku odlišnými metodami (triangulace),
- zjištění podobností a odlišností ve výsledcích při použití různých metod (komplementarita),
- použití jedné metody je nezbytnou podmínkou použití metody druhé (sekvenčnost),
- odhalení nových skutečností pro formulaci nebo reformulaci výzkumných otázek (iniciace) a

- rozšíření studie pomocí provedení více metod (expanze).

Souhrnné charakteristiky všech tří přístupu se nacházejí v Tabulce 1.

Tabulka 1 – Charakteristiky kvantitativního, kvalitativního a smíšeného výzkumu

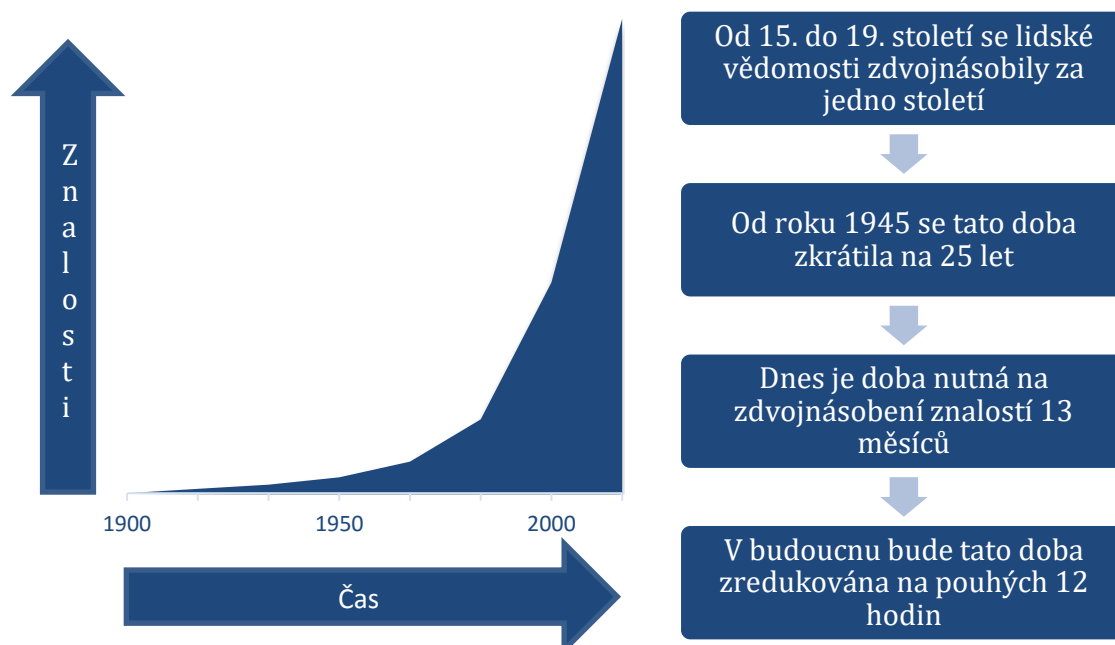
	Kvalitativní výzkum	Smíšený výzkum	Kvalitativní výzkum
vědecká metoda	deduktivní, testování hypotéz a teorie	deduktivní a induktivní	induktivní, generování teorie zakotvené v datech
pohled na chování člověka	určené faktory a má predikovatelný charakter	někdy predikovatelné	dynamické, situační, sociální, kontextuální
cíle výzkumu	popis, explorace a predikce	mnoho cílů	popis, explorace, objevování
zaměření	úzká perspektiva, testování hypotéz	více perspektiv	široká a hloubková perspektiva zacílená na hloubku a šířku případu
povaha pozorování	pokus zkoumat chování za kontrolovaných podmínek	zkoumání chování ve více kontextech	zkoumání chování v přirozených každodenních podmínkách
povaha reality	objektivní, různé pozorovatelé se shodnou na tom, co pozorují, pozitivistický přístup	realismus a pragmatismus	subjektivní, osobní a sociálně konstruované (také realismus)
forma dat	strukturovaná, validizovaná data, objektivní instrumenty měření	mnoho forem	sběr kvalitativních dat výzkumníkem (kvalitativní rozhovor, pozorování, dokumenty)
povaha dat	proměnné	směs proměnných, slov a obrazů	slova, obrazy, kategorie
analýza dat	identifikace statistických vztahů	kvantitativní nebo kvalitativní	hledání vzorců, témat a holistických vlastností
výsledky	zobecněné výsledky	směs různých výsledků	pohledy účastníků, návrh teorie, zdůrazněná mnohost perspektiv
výzkumná zpráva	statistická zpráva, korelace, predikční rovnice	eklektická, pragmatická	narativní zpráva s kontextuálními popisy a přímými citacemi výpovědi

Zdroj: Hendl (2005), vlastní zpracování

2 Statistika

S postupným rozšiřováním a zdokonalováním informačních technologií se rozšiřují také informace a znalosti, kterými lidstvo disponuje. Podle Schillinga (2013) se dnes lidské znalosti zdvojnásobují každých 13 měsíců, přitom na první zdvojnásobení znalostí si lidstvo muselo od roku 1 gregoriánského kalendáře počkat až do roku 1500. Na základě výzkumu provedeného pracovníky IBM se s rozšířením internetu větší očekává snížení této doby na pouhých 12 hodin.

Obrázek 3 – Vývoj znalostí



Zdroj: Fuller (1981), vlastní zpracování

Za exponenciálním růstem znalostí stojí ještě větší množství dat, ze kterých je nutné tyto znalosti prostřednictvím informací získat. Analyzování a vyhodnocování těchto a ostatních dat se zabývá statistika.

Statistika jako samostatná vědní disciplína vznikla až v novověku v 17. století. Její prvky se však využívaly již ve starověku. V této době lze základy statistiky objevit v záznamech o společenském a hospodářském životě, mezi které patřila evidence počtu osob či dobytka a úrody. Starověké říše, které byly velmi závislé na výběru daní, vedly velmi podrobné spisy o počtu obyvatelstva a jejich majetku. K obdobným účelům se prvky popisné statistiky využívaly i po celé období středověku a v raném novověku (Kačerová a Michalec, 2014).

Systematická analýza jevů a událostí na základě numerických záznamů se objevila v 17. století a je spojována s Angličany Johnem Grauntem a Williamem Pettym. Samotné slovo statistika bylo potom použito v roce 1749 na přednášce německého filosofa a statistika Gottfrieda Achenwalla a k tomuto datu se váže i vznik statistiky jako samostatného vědního oboru. Ve 20. a dále ve 21. století získala statistika nezapustitelnou pozici a lze tvrdit, že „v současnosti již neexistuje vědní obor, který by nepracoval s hromadnými daty a nevyužíval k jejich vyhodnocování statistiku (Hendl, 2005, s. 12)“.

2.1 Statistická analýza dat

Každé analýze dat musí bezpodmínečně předcházet jejich získání. Na počátku každého výzkumu se musí výzkumník rozhodnout, zda použije data sekundární nebo primární. Pojem sekundární data označuje data, která byla již shromážděna někým jiným za odlišným účelem. Výhoda těchto dat spočívá zejména ve finanční a časové úspoře potřebné k jejich získání. Naopak nevýhoda plyne ze skutečnosti, že byla shromážděna za účelem řešení jiného problému a tudíž nemusí plně vyhovovat jiným projektům (Kozel et al., 2011).

Pokud není možné získat relevantní sekundární data, je nezbytné shromáždit data primární. Problematika průzkumů a sběru dat byla již blíže nastíněna v předchozí kapitole a konkrétní technika sběru dat pro účely zpracování diplomové práce bude specifikována v metodologické části práce. Tato část bude proto zaměřena na analýzu již získaných dat.

Kvalitně provedená analýza získaných dat je klíčovým prvkem každého průzkumu a je podmínkou pro vyvození pravdivých závěrů. V dnešní době má výzkumník k dispozici širokou škálu nástrojů a možností, které může využít pro statistické vyhodnocení získaných dat. Z tohoto důvodu zmíním pouze ty nástroje, které mě i budoucímu čtenáři práce pomohou ke správnému pochopení problematiky a vyvození správných závěrů z nasbíraných dat. Mezi oblasti statistiky, ze kterých budu čerpat, patří problematika analýzy jednotlivých proměnných a testování hypotéz.

2.1.1 Analýza jednotlivých proměnných

Jednorozměrná statistická analýza je specifická tím, že se zabývá analýzou pouze jednoho statistického znaku bez ohledu, zda je součástí jednorozměrného nebo dvourozměrného statistického souboru.

Rozdělení četností

Rozdělení četností patří mezi elementární metody zpracování dat a slouží k zachycení absolutního nebo relativního zastoupení získaných hodnot na celkovém

souboru. Nejčastěji se tato metoda využívá pro analýzu kvalitativních proměnných. Samotná četnost lze definovat jako počet prvků s identickou hodnotou statistického znaku. Jednotlivé četnosti je možné zachytit pomocí tabulek nebo grafů (Řezanková, 2010).

V Tabulce 2 je zobrazeno základní schéma rozdělení četností. V této tabulce jsou jednotlivé kategorie proměnné x označeny písmenem x_i , přičemž $i = 1, 2, \dots, k$. K těmto hodnotám jsou vždy přiřazeny četnosti výskytu daného znaku n_i , kde $i = 1, 2, \dots, k$. V případě dotazníkového šetření bude hodnota n_i reprezentovat počet respondentů, kteří odpověděli totožně (Řezanková, 2010).

Tabulka 2 – Schéma rozdělení četností

Varianta znaku x_i	Četnost		Kumulativní četnost	
	absolutní n_i	relativní p_i	absolutní	relativní
x_1	n_1	p_1	n_1	p_1
x_2	n_2	p_2	$n_1 + n_2$	$p_1 + p_2$
...	...	...	...	...
x_k	n_k	p_k	$\sum_{i=1}^k n_i = 1$	$\sum_{i=1}^k p_i = 1$
Celkem	$\sum_{i=1}^k n_i = n$	$\sum_{i=1}^k p_i = 1$	x	x

Zdroj: Hindls et al. (2002), vlastní zpracování

V tabulkách lze hodnoty četností uvádět ve dvou tvarech, absolutním a relativním. Absolutní četnost vyjadřuje, kolikrát se hodnota daného znaku vyskytuje ve statistickém souboru. Pro účely porovnávání jednotlivých kategorií a konečnou interpretaci je však rychlejší a snazší využít ukazatele relativní četnosti. Ten udává podíl dané kategorie na celkovém souboru. Velmi často je tento ukazatel násoben 100 a vyjádřen v procentech.

V některých případech je základní tabulka četností rozšířena o kumulativní četnosti. Na jejich základě lze porovnávat kolik jednotek nebo jaká „část souboru má variantu znaku menší nebo rovnou určité dané obměně (Hindls, 2002, s. 18)“.

V oblasti problematiky četností je možné se setkat také s pojmem modální četnost. Ta vyjadřuje četnost nejvíce zastoupené hodnoty.

Kromě tabulek jsou vhodným prostředkem pro zobrazení četností také grafy. Jejich výhoda tkví zejména v jednoduchém a přehledném zobrazení analyzovaných skutečností. Z jednoduchosti plyne také zásadní nevýhoda grafického zobrazení,

která spočívá v možnosti jeho zneužití pro záměrnou manipulaci. Mezi nejčastěji používané typy grafů pro účely zobrazení četností patří sloupcový a výsečový graf. V případě intervalového rozdělení četností je nejvhodnější využít histogram.

Pomocí intervalových četností je možné popsat také numerická data. V takovém případě je však nutné nejprve určit optimální počet a velikost tříd nebo intervalů. Touto otázkou se zabývalo velké množství statistiků a v dnešní době se kromě subjektivního vyjádření využívají další dva přístupy, jejichž volba závisí na charakteristikách statistického znaku (Bartošová, 2006).

Prvním přístupem je Sturgesovo pravidlo a jeho použití je vhodné spíše pro menší datové soubory. Velikost jednotlivých intervalů se stanoví podle vztahu uvedeného níže. Zde k vyjadřuje počet tříd a n rozsah souboru:

$$k = 1 + 3,3 \log_{10} n$$

Pokud data vykazují normální rozdělení je možné počet tříd určit pomocí Scottova pravidla. Počet tříd je zde dán následujícím vzorcem:

$$k = \frac{n}{h_s}$$

Parametr h_s je v této rovnici určen vztahem:

$$h_s = \frac{3,5\sigma}{\sqrt[3]{n}}$$

Popisná statistika

Rozdělení četností je velmi užitečným nástrojem pro základní analýzu statistického souboru. Jeho použitelnost však není pro všechny statistické soubory vhodná a například porovnávání dvou statistických souborů pouze za pomoci četností by bylo velmi obtížné. Z těchto důvodů je nutné shrnout charakteristiku statistických souborů do několika málo ukazatelů, které jsou však dostatečně výstižné. Při této činnosti jsou využívány zejména míry polohy a míry variability, méně časté je využití doplňujících ukazatelů šikmosti a špičatost (Hindls et al., 2002).

Míry polohy (úrovně) zobecňují hodnoty statistického souboru pomocí jediné hodnoty. Mezi tyto ukazatele se kromě prostého minima a maxima řadí také různé typy středních hodnot a kvantily. Z pokročilejších a více robustních charakteristik patří do této skupiny například BES a Gastwirthův odhad (Bartošová, 2006).

Střední hodnoty v různých variacích jsou velmi hojně využívány pro shrnutí výsledků statistických šetření a nelze říci, že by některá ze středních hodnot byla ta jediná správná. Důležité je správné rozhodnutí, kterou ze středních hodnot pro získání konkrétní informace použít. Jednotlivé typy středních hodnot je možné rozdělit do dvou skupin podle toho, zda při jejich konstrukci dojde k využití všech jednotek

základního statistického souboru či nikoli. Podle tohoto hlediska se dělí na průměry, které se počítají z celého souboru a druhou skupinu, jejímiž nejvýznamnějšími zástupci jsou modus a medián (Hindls et al., 2002).

Velmi významným a pravděpodobně nejčastěji využívaným statistickým ukazatelem je aritmetický průměr. Jeho přednosti plynou z velmi jednoduché konstrukce a jasného významu. Z těchto důvodů se tento ukazatel hojně využívá také v běžném životě. Nevýhoda tohoto ukazatele plyne z jeho citlivosti na extrémní hodnoty, kdy je jediná odlehlá hodnota schopna zkreslit celý ukazatel. Prostý aritmetický průměr je možné vypočítat následujícím způsobem:

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

V některých případech je nutné zohlednit u jednotlivých variant znaku také jejich důležitost nebo váhu. Tento postup se využívá například v případech, kdy jsou jednotlivé hodnoty uspořádané v tabulce četností. Zde se pro určení váhy hodnot využívá jejich relativní četnost. Ve výsledném vzorci je potom každá hodnota násobena svou váhou w_i .

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

Kromě aritmetického průměru se také využívá průměr harmonický. Ten není ve statistické praxi příliš využíván a své uplatnění našel zejména v indexové teorii. Tento průměr „je definován jako podíl počtu pozorování a součtu převrácených hodnot znaku (Hindls, 2002, s. 32)“ a stejně jako u aritmetického průměru existuje také jeho vážená varianta nazývaná vážený harmonický průměr.

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

Další alternativou je geometrický průměr, který se využívá při analýze vývoje ukazatele v čase například pro výpočet průměrného tempa růstu určitého ukazatele.

$$\bar{x}_G = \sqrt[n]{x_1 x_2 \dots x_n}$$

Posledním častěji využívaným průměrovým ukazatelem je kvadratický průměr a jeho vážená varianta. Spíše než pro ekonomickou praxi je tento ukazatel vhodný pro studijní účely, kdy jeho pochopení velmi zjednoduší pochopení dalších statistických ukazatelů.

$$\bar{x}_K = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$$

Mezi jednotlivými typy průměrů panuje následující vztah:

$$\bar{x}_H \leq \bar{x}_G \leq \bar{x} \leq \bar{x}_K$$

Znaménko rovnosti platí pouze v případě, že jsou všechny varianty znaku x stejné. Tato varianta je však velmi vzácná, protože pokud výzkumník zjistí, že jsou všechny znaky stejné, nemá důvod využít průměrové ukazatele. V situacích, kdy se alespoň jedna hodnota odlišuje, platí relace nerovnosti dle vztahu výše (Hindls et al., 2002).

Z problematiky kvantilů je nezbytné zmínit pojem medián. Tento 50% kvantil rozděluje statistický soubor na dvě stejně velké poloviny. To znamená, že polovina hodnot uspořádaných vzestupně je menší než tento ukazatel a druhá polovina nabývá hodnot vyšších. Při lichém počtu znaků je mediánem prostřední hodnota. Pokud je rozsah souboru sudý, je medián stanoven jako aritmetický průměr dvou hodnot ležících uprostřed souboru. Hlavní devizou tohoto ukazatele je jeho necitlivost na extrémní či odlehlé hodnoty, které se v souboru mohou vyskytovat (Hindls et al., 2002).

Pro správnou interpretaci statistických výsledků je důležité pochopení vztahu mezi aritmetickým průměrem a mediánem. Rovnost či nerovnost těchto ukazatelů závisí na šikmosti rozdělení hodnot, proto se průměr a medián rovnají pouze v případě symetrického rozdělení.

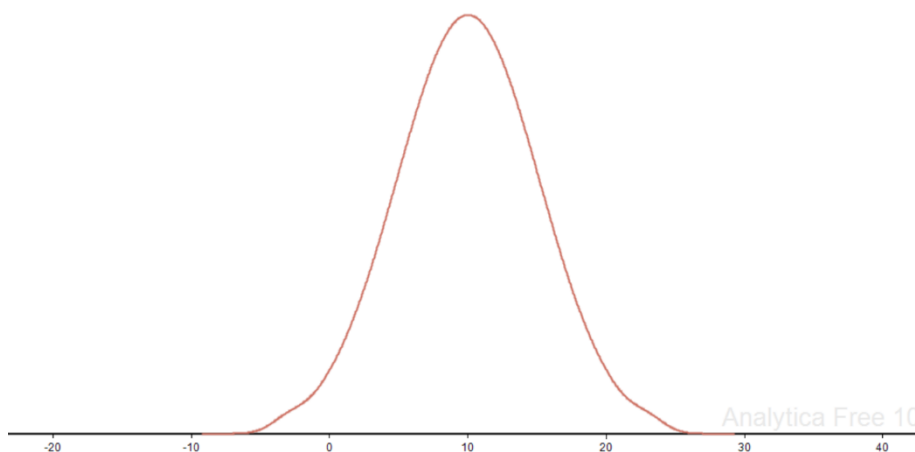
Velmi diskutované je využití mediánu při analýze rozložení mzdy ve společnosti. V České republice je podle Českého statistického úřadu ve 3. čtvrtletí roku 2016 průměrná mzda 27 220 Kč, přičemž medián tohoto ukazatele je 23 527 Kč. Z porovnání těchto ukazatelů plyne, že na průměrnou mzdu nedosahuje polovina Čechů (Český statistický úřad, 2016).

Dalšími hojněji využívanými kvantilovými ukazateli jsou kvartily a decily. Kvartily rozdělují vzestupně uspořádaný statistický soubor na 4 stejně početné části. Podobnou funkci plní decily, které však dělí statistický soubor na 10 částí. V osobním životě se lze s kvantily setkat například při vyhodnocování přijímacích testů, jako ukazatelem úspěšnosti studenta mezi ostatními.

Posledním ukazatelem z této oblasti je modus, který reprezentuje nejčastěji se vyskytující hodnotu. Pokud je v souboru pouze jeden mód, je tento soubor nazýván unimodálním. Módů se může ve statistickém souboru vyskytovat i více, pokud se některé hodnoty objevují stejně často. V případě existence dvou módů se jedná o bimodální soubor (Řezanková, 2010).

Prozatím shrnuté ukazatele popisné statistiky poskytují informace pouze o charakteristice polohy či úrovně. Velmi častým jevem jsou však případy, ve kterých se soubory se shodnou polohou budou výrazně lišit. Příkladem je Obrázek 4 a Obrázek 5, které zobrazují soubory s normálním rozdělením a se shodným aritmetickým průměrem. Přesto je možné pozorovat u obou grafů velmi markantní odlišnosti.

Obrázek 4 – Normální rozdělení s průměrem 10



Zdroj: vlastní zpracování

Obrázek 5 – Normální rozdělení s průměrem 10



Zdroj: vlastní zpracování

Hodnoty souboru zobrazeného na Obrázku 4 jsou více koncentrovány okolo průměru. Je tedy možné říci, že průměr má v tomto případě lepší vypovídací schopnost než v případě rozdělení na Obrázku 5. Míry variability nebo rozptýlenosti se tedy snaží zachytit rozptýlenost hodnot statistického souboru kolem stanovené střední hodnoty.

Jak již bylo ukázáno na předešlém příkladu, jednou z funkcí **charakteristik variability** je zhodnocení vypovídající hodnoty aritmetického průměru. Obecně lze konstatovat, že „čím je odlišnost sledovaného znaku větší, tím menší má průměr vypovídací schopnost (Bartošová, 2006, s. 27)“.

Posouzení vypovídací schopnosti aritmetického průměru však není jedinou funkcí měr variability. Podle Hindlse (2002) jsou tyto charakteristiky jedněmi z nejdůležitějších v celé statistice. Toto tvrzení založil na skutečnosti, že v řadě statistických disciplín je nutné správně definovat proměnlivost obměn statistických znaků.

Jednotlivé ukazatele je možné rozdělit do dvou skupin na míry absolutní variability a míry relativní variability.

Míry absolutní variability se od druhé skupiny liší tím, že „vyjadřují variabilitu ve stejných měrových jednotkách, v nichž je vyjadřován sledovaný znak (Hindls, 2002, s. 35)“.

Základním ukazatelem absolutní variability je variační rozpětí. Tento ukazatel vyjadřuje rozdíl mezi nejvyšší a nejnižší hodnotou.

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

Výpočet hodnoty variačního rozpětí je stejně jako samotná interpretace výsledku velmi jednoduchá. Z této jednoduchosti však plyne i nevýhoda, jelikož variační rozpětí nepodává informaci o heterogenitě hodnot uvnitř rozpětí.

Stejně jako průměr je ukazatel variačního rozpětí velmi citlivý na extrémní hodnoty, které mohou značně ovlivnit výsledek. Z důvodu eliminace odlehlých hodnot je používáno kvartilové rozpětí, které představuje rozdíl mezi horním a dolním kvartilem.

$$R_q = \tilde{x}_{0,75} - \tilde{x}_{0,25}$$

Vyšší vypovídající schopnost mají takové ukazatele variability, jejichž hodnota je závislá na heterogenitě všech hodnot souboru. Nejpoužívanějším ukazatelem variability je rozptyl. Ten je definován následujícím vzorcem:

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Rozptyl tedy vyjadřuje disperzi hodnot souboru kolem aritmetického průměru. Čím vyšší variabilita hodnot, tím je vyšší ukazatel rozptylu. Pokud například budou všechny hodnoty souboru totožné, bude rozptyl tohoto výběru roven nule (Hindls et al., 2002).

Ukazatel rozptylu je vzhledem k jednotkám základního souboru uváděn vždy v kvadratických jednotkách. Toto vyjádření může komplikovat následnou interpretaci, proto je používán ukazatel směrodatné odchylky. Z pohledu výpočtu se jedná o druhou odmocninu z rozptylu.

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

Míry relativní variability nejsou vyjádřeny ve stejných jednotkách jako hodnoty analyzovaného souboru. Důvodem je způsob výpočtu, jelikož se při konstrukci těchto ukazatelů využívají ukazatele absolutní variability, které se dávají do poměru k průměru nebo mediánu. Díky tomu lze porovnávat variabilitu znaků, které nemají shodné jednotky nebo které se liší mírou polohy.

Hlavním ukazatelem relativní variability je variační koeficient, který vyjadřuje poměr směrodatné odchylky a aritmetického průměru.

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

Velmi často se tento ukazatel objevuje v procentech, přičemž pokud dosahuje hodnot vyšších než 50 %, jedná se o nesourodý statistický soubor.

Druhým ukazatelem relativní variability je relativní kvartilová odchylka. Její význam ve všech případech v porovnání s variačním koeficientem malý.

$$Q_{rel} = \frac{\tilde{x}_{75} - \tilde{x}_{25}}{\tilde{x}_{75} + \tilde{x}_{25}}$$

Mezi poslední častěji využívané ukazatele popisné statistiky patří šikmost a špičatost. Tyto ukazatele se zabývají především charakteristikou rozložení hodnot ve statistickém souboru. Podle koeficientu šikmosti lze posuzovat, zda jsou jednotlivé hodnoty statistického souboru symetricky rozloženy kolem střední hodnoty nebo inklinují na jednu stranu (Kozel et al., 2011).

Koeficient špičatosti vyjadřuje stupeň koncentrace hodnot souboru okolo střední hodnoty. Pro Gaussovo rozdělení nabývá tento koeficient hodnoty 0. Pokud jsou hodnoty nižší než 0, nejsou hodnoty koncentrovány okolo střední hodnoty. V případě kladných hodnot jsou hodnoty znaku rozmístěny rovnoměrně (Fotr a Hnilica, 2014).

2.1.2 Testování hypotéz

Ve statistice má testování hypotéz velmi významnou pozici. Swoboda (1977) dokonce tvrdí, že moderní statistika je z velké části tvořena testováním hypotéz. Hypotézy lze charakterizovat jako jednoduše formulované předpoklady o rozdělení zkoumaného znaku nebo o vazbách mezi jednotlivými proměnnými.

Zdrojů pro formulaci hypotéz může být hned několik. Kozel et al. (2011) rozlišuje následujících 5 nejčastějších zdrojů:

- „dřívější praktické zkušenosti,
- teoretické znalosti,
- dostupné statistické databáze,
- explorativní výzkum a
- přání zadavatele výzkumu (Kozel et al., 2011, s. 78)“.

Na počátku každého testování je stanovení hypotéz, které budou testu podrobeny. Jako první je vždy nulové hypotéza H_0 , která je někdy označována jako testovaná. Tato hypotéza vyjadřuje určitou shodu, rovnost či nezávislost proměnných. Typickým příkladem tohoto druhu hypotéz může být následující tvrzení:

„Nová Škoda Octavia je stejně kvalitní jako starý model.“

Proti nulové hypotéze je postavena tzv. hypotéza alternativní H_1 . Ta vždy specifickým způsobem popírá platnost hypotézy nulové. K předchozímu příkladu je možné stanovit následující alternativní hypotézu:

„Nová Škoda Octavia je kvalitnější než starý model.“

Z uvedených příkladů vyplývá, že jednotlivé hypotézy si navzájem odporují.

Většina statistických šetření nemá k dispozici hodnoty pro všechny jednotky základního statistického souboru. Testování hypotéz proto probíhá pouze na určité části souboru, který je nazýván výběrovým souborem. Z tohoto důvodu je možné, že při testování dojde vyvození chybných závěrů a následně k zamítnutí platné hypotézy.

Při zamítnutí hypotézy H_0 , která ve skutečnosti platí, se výzkumník dopouští chyby prvního druhu. Pravděpodobnost této chyby se značí řeckým písmenem α . Tato chyba vzniká tím, že dojde k odmítnutí dobré nebo odpovídající hypotézy na základě nesprávného vzorku (Swoboda, 1977).

Stejně tak mohou nastat případy, kdy výzkumník přijme špatnou nebo neodpovídající hypotézu na základě příznivého vzorku. To znamená přijetí hypotézy H_0 , přestože ve skutečnosti platí hypotéza H_1 . V takové situaci hovoříme o chybu druhého druhu, jejíž pravděpodobnost nese označení β (Hindls et al., 2002).

Tabulka 3 obsahuje jednoduché schéma, ze kterého je patrný vztah chyb obou druhů.

Tabulka 3 – Chyby I. a II. druhu

Rozhodnutí o H_0	Skutečnost	
	H_0 pravdivá	H_1 pravdivá
Nezamítnutí	správné rozhodnutí	chyba II. druhu
Zamítnutí	chyba I. druhu	správné rozhodnutí

Zdroj: Kozel et al. (2011), vlastní zpracování

V Tabulce 4 jsou uvedeny pravděpodobnosti, se kterou chyby obou druhů nastávají.

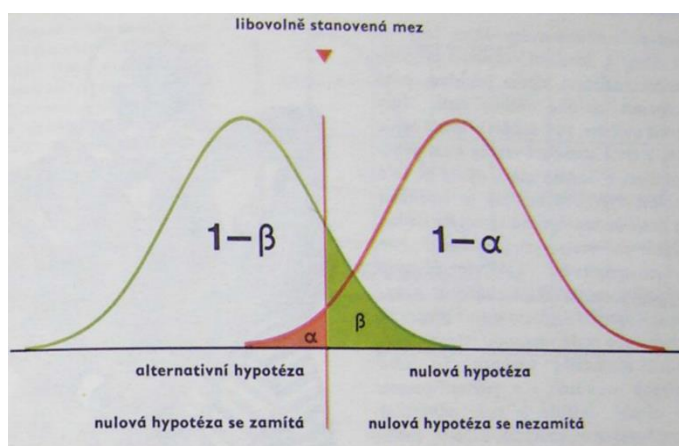
Tabulka 4 – Pravděpodobnosti chyb I. a II. druhu

Rozhodnutí o H_0	Pravděpodobnost	
	H_0 pravdivá	H_1 pravdivá
Nezamítnutí	$1-\alpha$	β
Zamítnutí	α	$1-\beta$

Zdroj: Hindls et al. (2002), vlastní zpracování

Pravděpodobnost α se nazývá hladina významnosti. Tato hladina vyjadřuje maximální přípustnou toleranci chyby I. druhu. Hodnota této pravděpodobnosti je určena předem a obvykle je 0,05. V případech, kdy je vysoký tlak na redukci této chyby a je k dispozici široký rozsah výběru, se využívá také 0,01 či 0,01. Je však nutné zmínit, že při snižování této hodnoty dochází ke zvyšování pravděpodobnosti chyby II. druhu (viz Obrázek 6). Tento efekt je umocněn v situacích, kdy výzkumník nemá k dispozici dostatečné množství dat (Hindls et al., 2002).

Obrázek 6 – Vztah chyby I. a II. druhu



Zdroj: Swoboda (1977)

Druhým významným ukazatelem Tabulky 4 je síla testu, která má označení $1 - \beta$. Ta udává pravděpodobnost, že nedojde k chybě II. druhu. Tedy, že nebude přijata neplatná hypotéza H_0 na úkor platné hypotézy H_1 .

Díky rozvoji statistických programů je testování hypotéz velmi pohodlné. O zamítnutí či nezamítnutí hypotézy H_0 se výzkumník rozhoduje na základě p-hodnoty, kterou dokáží statistické programy sami spočítat. Výsledná p-hodnota se následně porovnává s hladinou významnosti α a na základě jejich vztahu dochází k rozhodnutí o hypotézách (viz Tabulka 5).

Tabulka 5 – P-hodnota a hypotézy

$p \leq \alpha$	• zamítnutí H_0 • prokázání H_1
$p > \alpha$	• nezamítnutí H_0 • neprokázání H_1

Zdroj: Swoboda (1977), vlastní zpracování

3 Metodologie diplomové práce

Hlavním cílem práce je nalezení odpovědí na výzkumné otázky, které jsou definované v úvodu diplomové práce. Podmínkou pro naplnění stanovených cílů je správné použití vhodných metod. Jelikož zpracovávané téma je z oblasti statistiky, je nutné získat určitá data, která je možno dále analyzovat a vyvozovat potřebné závěry. Pro získání dat bude použita jedna z metod primárního sběru dat, a to dotazníkové šetření. Pro účely vyvrácení nebo potvrzení stanovených hypotéz bude aplikována metoda Personova Chí – kvadrátu testu nezávislosti.

3.1 Dotazníkové šetření

Dotazníkové šetření je jednou z nejčastěji používaných metod kvantitativního průzkumu. Výhoda dotazníku spočívá v tom, že jeho prostřednictvím je možné postihnout vysoký počet dotazovaných za krátký časový úsek a při investování nízkých nákladů.

Volba konkrétního typu dotazníku je ovlivněna řadou faktorů. Podle Saundorse (2009) závisí volba konkrétního dotazníku na:

- charakteristice respondentů, jež mají být dotazováni podrobeni,
- počtu otázek, které mají být v rámci dotazování kladeny,
- délce a komplikovanosti otázek,
- finančních možnostech,
- časové náročnosti sběru dat.

Velmi významným rozhodnutím, které musí zhotovitel udělat před konstrukcí samotného dotazníku, je volba typů otázek podle variant odpovědí. Kozel et al. (2011) uvádí celkem 3 různé typy: otevřené, uzavřené a polouzavřené.

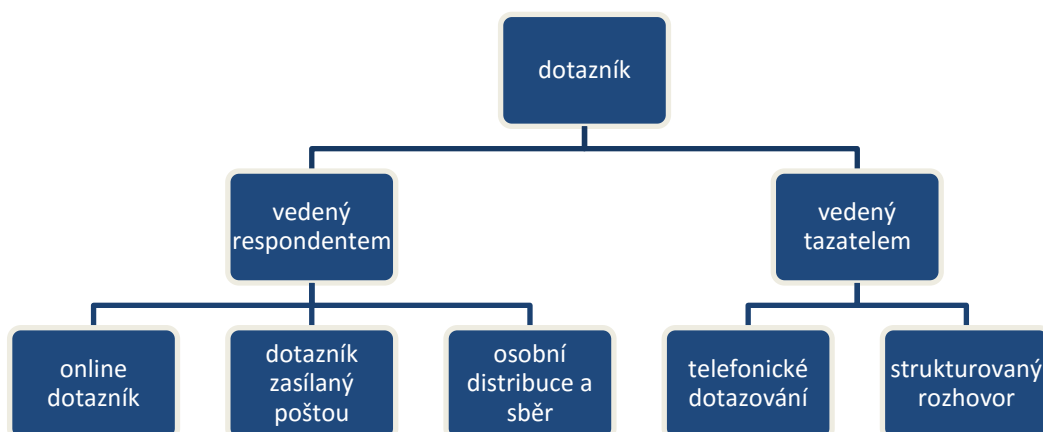
Pokud je dotazník nebo jeho část sestavena z otevřených otázek, nejsou respondenti nabízeny žádné varianty odpovědí. Respondent tudíž není žádným způsobem omezován ve svých výpovědích. Tazatel díky tomu získá celou řadu jedinečných názorů, jejich následná analýza však bývá velmi obtížná. Často je tento typ otázek využíván pro kvalitativní výzkum.

U uzavřené varianty otázek je respondentovi nabídnuto různé množství variant odpovědí. Tato varianta nachází uplatnění zejména u kvantitativních průzkumů. Díky standardizaci odpovědí spočívají přednosti této metody v rychlosti jak při sběru, tak i následné analýze dat.

Kompromis mezi oběma zmíněnými variantami je reprezentován polouzavřenými otázkami. Dotazovaný může zvolit jednu z předložených odpovědí, ale také je mu poskytnuto právo napsat odpověď libovolnou.

Pokud tazatel vezme v úvahu výše uvedené faktory, může přistoupit k volbě konkrétního typu dotazníku. Saunders (2009) rozlišuje dva základní druhy dotazníků podle kontaktu mezi respondentem a tazatelem (viz Obrázek 7).

Obrázek 7 – Typologie dotazníků



Zdroj: Saunders et al. (2009), vlastní překlad a zpracování

Pro výzkum prováděný v rámci diplomové práce bude použita metoda uzavřených otázek prostřednictvím online dotazníku, doplněná osobní distribucí a sběrem.

Metodu online dotazování jsem vybral zejména díky její schopnosti postihnout vysoké množství respondentů z různých částí České republiky při vynaložení nízkých nákladů. Naopak nevýhodu tohoto druhu dotazování spatřuji ve skutečnosti, že již na začátku šetření z průzkumu vylučuji osoby, které nemají k dispozici připojení na internet. Tento nedostatek se projeví především vyloučením starší části populace a sociálně slabších skupin. K eliminaci této nevýhody doplním online metodu dotazníkového šetření o osobní distribuci a sběr dotazníků. Tím budou do vzorku populace zahrnuty i osoby, které online dotazník není schopen postihnout.

Samotný dotazník bude strukturovaný do třech relativně samostatných oddílů.

První oddíl bude zaměřen na ověření znalosti statistických pojmů. Zde bude sledována především schopnost respondenta vybrat z předložených odpovědí tu správnou pro daný pojem. Do dotazníkového šetření budou zařazeny pojmy jako průměr, modus, medián nebo rozptyl.

Ve druhé části dotazníku budou respondentovi předkládány zprávy nebo grafy, z nichž některé budou účelně konstruované. Úkolem respondenta bude z daných sdělení vyvodit správné závěry. Z výsledků tohoto testu bude možné určit, zda jsou dané osoby schopny si i přes účelné zkreslení odnést relevantní informace.

Třetí oddíl bude zjišťovat základní charakteristiky respondentů zapojených do dotazníkového šetření. Tím budou získány údaje o pohlaví, věku, vzdělání a kraji, ve kterém daný respondent bydlí.

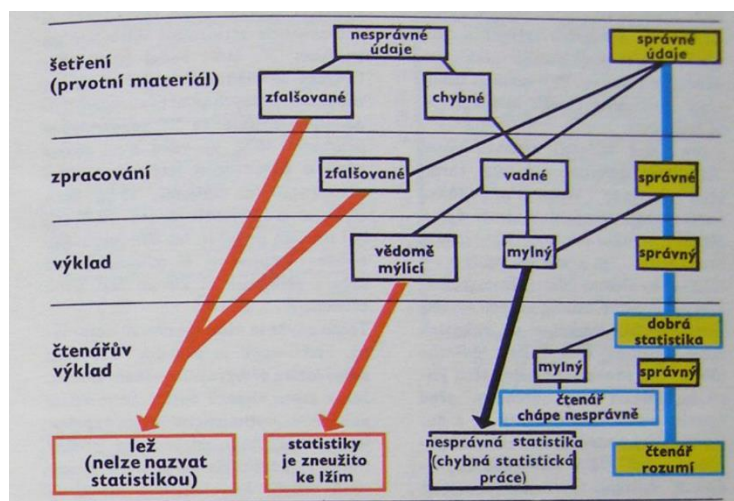
4 Jak „lhát“ se statistikou

Americký spisovatel Mark Twain ve své autobiografické knize nazvané „*Mark Twain's Own Autobiography: The Chapters from the North American Review*“ zmiňuje pravděpodobně nejznámější citát týkající se statistiky, a to že existují pouze tři druhy lži: lži, zatracené lži a statistika. Autorství tohoto citátu bylo po desetiletí připisováno bývalému premiéru Spojeného království Benjaminu Disraelimu. Dnes je podle posledních studií za původce tohoto citátu považován britský politik Leonard H. Courtney (The University of York, 2012).

Na druhé straně však existuje postoj, který statistiku nebo statistické výsledky považuje za nevyvratitelné vyjádření přesnosti. Velké množství autorů používá k zajištění nezpochybnitelnosti nebo pocitu nezpochybnitelnosti svých tvrzení fráze typu „*statistické výsledky dokazují, na základě statistických šetření*“ apod. Takto protichůdné náhledy na výsledky statistiky pak vzbuzují v laickovi zmatek, který přirozeně vede k tomu, že se pro něj statistika stává velmi záhadnou vědou. Podle Swobody (1977, s. 18) za takto rozdílné pojetí statistiky mohou dvě skutečnosti: „*za prvé nedostatečná znalost cílů, metod a možností statistiky a za druhé, že za statistiku se pokládá i to, co je ve skutečnosti pseudostatistikou*“.

V této části práce se budu zabývat zejména prvním postojem ke statistice, kdy je tato vědní disciplína využívána k ovlivňování vnímání jednotlivce a posléze celé společnosti. K ovlivnění osob za využití statistiky či pseudostatistiky může docházet několika způsoby (Obrázek 8).

Obrázek 8 – Zdroje chyb a možnosti klamání



Zdroj: Swoboda (1977)

Jednotlivé podoby zmíněné v této diplomové práci vzešly syntézou poznatků z knih „How to lie with statistics“ od Derrella Huffa a „Moderní statistika“, jejímž autorem je Helmut Swoboda. Jednotlivé způsoby jsou také doplněny ukázkami využití těchto metod v praxi.

4.1 Problém nereprezentativního vzorku

Většina prováděných statistických šetření není založena na datech od všech statistických jednotek dané populace, ale snaží se závěry získané na základě vzorku populace zobecnit na celou populaci. Aby vybraný vzorek byl schopný reprezentovat danou populaci, musí mít všechny jednotky shodnou šanci dostat se do výběrového souboru. Pokud je však u některé z jednotek populace tato šance větší než u druhé, je náhodnost výběru narušena systematickou chybou.

Systematická chyba má za následek výběr nereprezentativního vzorku populace, který v konečném důsledku způsobí zobecnění neplatných závěrů pro celou populaci a nepřesnost průzkumu.

Jedním z nejznámějších případů, kdy k chybě způsobené výběrem nereprezentativního vzorku došlo, je průzkum výsledků voleb amerického prezidenta v roce 1936. V tomto roce rozeslal magazín Literary Digest téměř 10 milionů dotazníků, přičemž respondenti byli vybíráni z telefonních seznamů a seznamů držitelů řidičského oprávnění. Celkem se vrátilo přes 2,3 milionu dotazníků. Toto na první pohled oslnivé číslo však znamená účast menší než 25 %. Výsledky tohoto šetření vyzněly kladně pro republikánského vyzyvatele Alfa Landona, který s 55 % porazil demokratického kandidáta Franklina Roosevelta o 14 % (Tabulka 6). Pro Roosevelta to byla již druhá prezidentská kandidatura.

Tabulka 6 – Výsledky průzkumu prezidentských voleb USA v roce 1936

Jméno kandidáta	Výsledky průzkumu	
	Počet hlasů	Počet hlasů v %
Franklin Roosevelt	972 897	41,40 %
Alf Landon	1 293 669	55,05 %
William Lemke	83 610	3,56 %

Zdroj: Squire (1988), vlastní zpracování

Výsledky voleb však dopadly naprosto odlišně a Franklin Roosevelt vyhrál o celých 24 % (Tabulka 7).

Tabulka 7 – Výsledky prezidentských voleb USA v roce 1936

Jméno kandidáta	Výsledky voleb	
	Počet hlasů	Počet hlasů v %
Franklin Roosevelt	27 752 648	60,80 %
Alf Landon	16 681 862	36,54 %
William Lemke	892 378	1,95 %
ostatní	320 811	0,70 %

Zdroj: Leip (2016), vlastní zpracování

Tato fatální chyba měla velký podíl na krachu magazínu v roce 1938 (Squire, 1988).

Časopis se dopustil zásadní chyby při výběr respondentů. Kvůli tomu, že do základního souboru zahrnul pouze majitele telefonů a automobilů, došlo k eliminaci nejnižší vrstvy obyvatel. Zajímavé je, že magazín se této chyby dopustil také ve volebních letech 1920, 1924, 1928 a 1932, a přesto byly výsledky přesné. Hlavním důvodem byla Velká deprese, která vrcholila v roce 1935 a způsobila výraznou polarizaci společnosti a donutila k účasti ve volbách v roce 1936 také voliče z nižších sociálních vrstev. Skutečnost, že průzkumy Literary Digest v letech 1920-1932 odpovídaly následným volbám, bylo pouze dílem náhody (Disman, 2011).

Volby zejména do vlivných funkcí jsou vždy velmi sledovaným společenským tématem. Především prezidentské volby v USA, které probíhaly 8. listopadu 2016, neunikly pozornosti většině agentur zabývajících se problematikou průzkumů. Oproti vzorku použitého v roce 1936 byl počet respondentů v roce 2016 pouze zlomkový (Tabulka 8). Ten však dokázal mnohem přesněji reflektovat skutečnost, kdy Hilary Clintonová získala v celkovém součtu větší množství hlasů, i když nakonec kvůli volebnímu systému USA neztvrdila.

Tabulka 8 – Výsledky průzkumu prezidentských voleb USA v roce 2016

Hlasování	Velikost vzorku	Clinton	Trump
Konečný výsledek	--	48,2 %	46,1 %
RCP Average	--	45,5 %	42,2 %
Bloomberg	799	44 %	41 %
IBD/TIPP Tracking	1 026	43 %	45 %
Economist/YouGov	3 677	45 %	41 %
ABC/Wash Post Tracking	2 220	47 %	43 %
FOX News	1 295	48 %	44 %
Monmouth	748	50 %	44 %
Gravis	16 639	47 %	43 %
NBC News/Wall St. Jrnl	1 282	44 %	40 %
Reuters/Ipsos	2 196	42 %	39 %
Rasmussen Reports	1 500	45 %	43 %
CBS News	1 426	45 %	41 %

Zdroj: Real Clear Politics (2016), vlastní zpracování

Je důležité si uvědomit, že chyba způsobená společností Literary Digest se může se stejnými důsledky objevit v kterémkoli šetření nebo anketě. Je tedy vždy nutné dávat pozor, na základě jakých dat a od jakých respondentů pocházejí dané závěry jednotlivých statistických šetření.

4.2 Zavádějící střední hodnoty

V části práce, která se týkala vymezení jednotlivých statistických pojmů, byla nejpochetněji zastoupena skupina středních hodnot. Existence mnoha metod, kterými je možné tyto skalární ukazatele sestavit, může v posluchači či čtenáři způsobit značný zmatek. Ze stejného důvodu také pramení riziko pro manipulaci s veřejným míněním, kdy autor účelně využívá určitý typ střední hodnoty.

Výborný příklad dokonalého zmatení čtenáře či posluchače prostřednictvím rozdílných podob středních hodnot uvádí Swoboda (1977). Muž rozmyšlí koupit pozemku ve městě Zbohatlíkov a zcela logicky se zajímá o roční příjmy obyvatel tohoto regionu. Od realitního makléře dostává informaci o hodnotě 82 320 dolarů, a že 20 % obyvatel tohoto města má průměrný příjem 308 400 dolarů. Bankovní zástupce uvedl, že více než polovina obyvatel má roční příjem přesahující 29 000 dolarů a nejčastější roční výdělek je 18 000 dolarů. Místní počtář, který se detailně zabýval místní

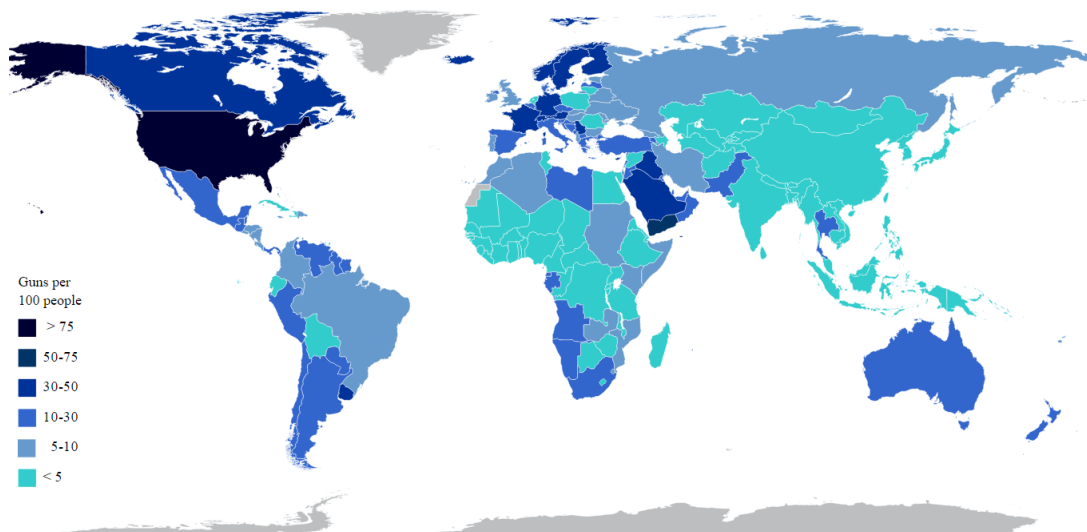
statistikou, uvedl, že většina obyvatel má příjem menší než 7 500 dolarů. Na první pohled nesmyslné informace o průměrných příjmech, které uvedli jednotliví pracovníci, jsou však opravdu pravdivé. V malém městečku bydlel jeden milionář a matematický postup pro konstrukci mediánu, módu a ostatních typů středních hodnot je diametrálně odlišný. Z tohoto důvodu je nutné, aby každý autor vždy uváděl, o kterou z hodnot se v jeho případě jedná.

Stejně jako v městečku Zbohatlíkov se střední hodnoty využívají k analýze rozložení příjmů ve společnosti i v reálném světě. V reálném světě je za tímto účelem nejčastěji používán aritmetický průměr. Jelikož se však v populaci vyskytují jednotlivci, jejichž příjem několikanásobně převyšuje příjem průměrný, dosáhne například v České republice na průměrný příjem pouze třetina obyvatel.

Aritmetický průměr je pro vyjádření rozložení příjmů ve společnosti opravdu zavádějící. Důkazem je také situace z amerického města Steubenville, ve kterém byl průměrný příjem v 7 878 domácnostech v roce 2011 46 341 \$. Pokud by se však do tohoto města přestěhoval Warren Buffett a Oprah Winfrey, přes noc by tato hodnota vzrostla o 62 % na 75 263 \$. Mediánové vyjádření tohoto ukazatele by však zůstalo, stejně jako reálná ekonomická situace obyvatel Steubenville, takže v nezměněné podobě (Coontz, 2013).

V roce 2007 vydal nezávislý projektový tým The Small Arms Survey studii s názvem „*Guns and the City*“. V této knize se objevují data o průměrném počtu zbraní na 100 obyvatel. Na prvním místě ze všech zemí světa se podle této studie umístily Spojené státy americké, kde na 100 rezidentů připadalo v roce 2007 89 zbraní (viz Obrázek 9). Tato zpráva se záhy objevila takže ve všech předních světových médiích jako například The Washington Post, CBC News nebo The Guardian.

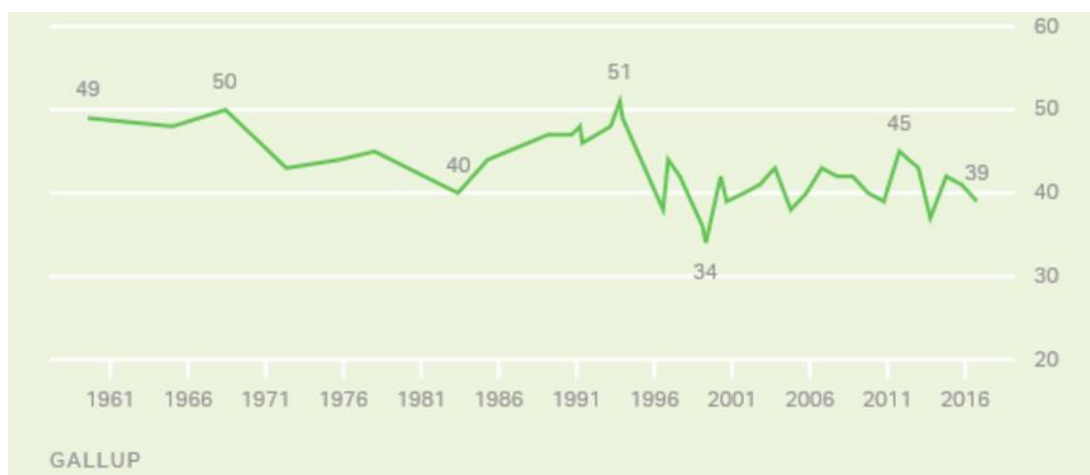
Obrázek 9 – Průměrný počet zbraní na 100 obyvatel



Zdroj: Small Arms Survey (2007)

I přesto, že se jedná o opravdu vysoké číslo, není možné na základě této informace tvrdit, že téměř každý občan Spojených států amerických je držitelem zbraně. Důležité je pohlížet na rozložení těchto zbraní ve společnosti. Takovou studii se zabývala průzkumná agentura The Gallup Poll. Na základě průzkumu z roku 2016 zjistila, že zbraň se nachází v 39 % amerických domácností (viz Obrázek 10). Podobného výsledku dosáhl také průzkum agentury PEW Research Center, podle kterého vlastní zbraň v roce 2016 44 % amerických domácností.

Obrázek 10 – Vývoj amerických domácností vlastnicí zbraně



Zdroj: Gallup (2016)

Příklad se zbraněmi uvedený výše dokonale reflektuje nedokonalosti průměru. Ten nepodává obraz o rozložení zbraní ve společnosti a několik majitelů velkého množství zbraní značně zkresluje konečný aritmetický průměr.

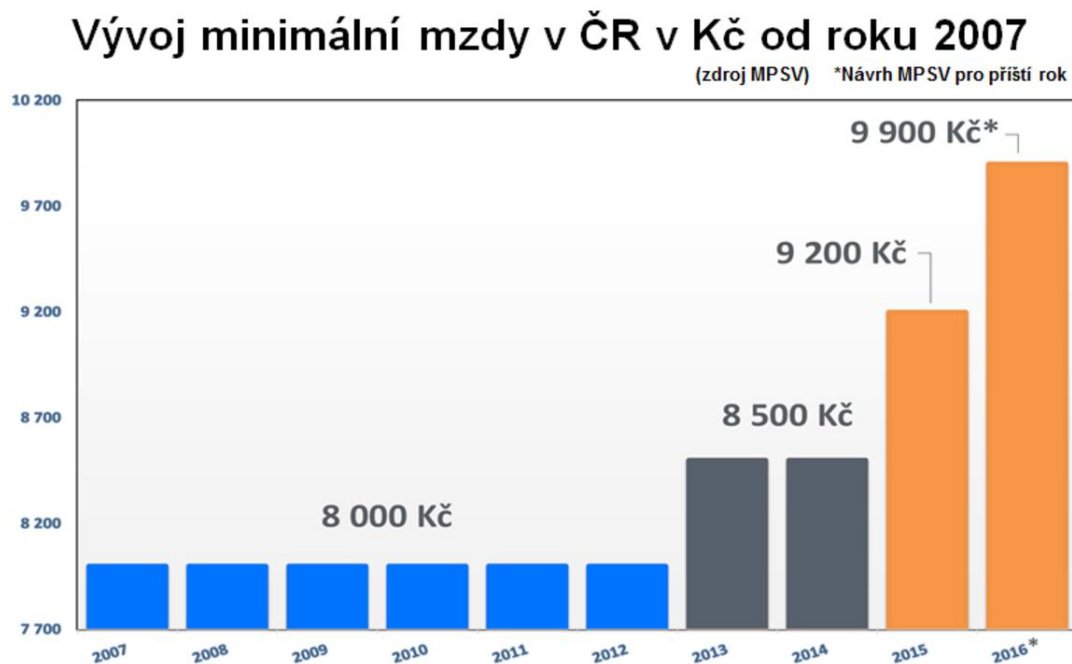
4.3 Změť grafů

Grafické zobrazení určitých statistických ukazatelů je v dnešní době velmi oblíbené a často doplňuje zprávy, prezentace či odborné texty. Důvody jsou jednoduché, grafické zobrazení usnadní porozumění a pochopení daného sdělení a výrazně oživí často nudné odstavce textu či nepřehledný soubor tabulek. Tato jednoduchost a popularita grafů však skýtá značné riziko zneužití. Pro manipulátora je totiž snadné vnutit příjemci prostřednictvím rafinovaně upraveného grafu velmi zkreslenou skutečnost a vyvodit v jeho mysli nesprávné závěry. Velmi výstižně tuto problematiku shrnuje Zamrazilová (2013, s. 88), která v komentáři ke knize *„Jak lhát se statistikou“* píše, že dnes *„dokážeme vytvořit graf, který skutečně ohromí a dokáže vyvolat tu představu, jakou chceme, i když původní surová data nám do karet moc nehrají. Posun měřítko osy, maxima či minima, může dokázat pravé divy“*.

Pravděpodobněji nejčastější chybou z této oblasti je konstrukce sloupcového či spojnicového grafu, ve kterém je záměrně zkrácen interval osy Y. Taková úprava grafu vyvolá v recipientovi informace dojem, že rozdíly mezi jednotlivými ukazateli jsou propastné. Ve skutečnosti jsou však minimální. V některých případech je však zúžení intervalu osy Y nutné. Jde zejména o trendy, které se v jednotlivých obdobích od sebe liší minimálně. Při zobrazení celého intervalu v grafu by tak výsledkem byla pouze rovná čára a trend by nebyl zřejmý. Výše zmíněný způsob manipulace se netýká pouze soukromého sektoru, nýbrž se objevuje i v nejvyšších kruzích české politické sféry.

Dne 18. května 2015 se za účelem jednání Rady hospodářské a sociální dohody konalo ve Strakově akademii zasedání tripartity. Úřad vlády České republiky na tomto setkání prezentoval vývoj minimální mzdy od roku 2007 (Obrázek 11).

Obrázek 11 – Graf prezentovaný na zasedání tripartity dne 18. května 2015

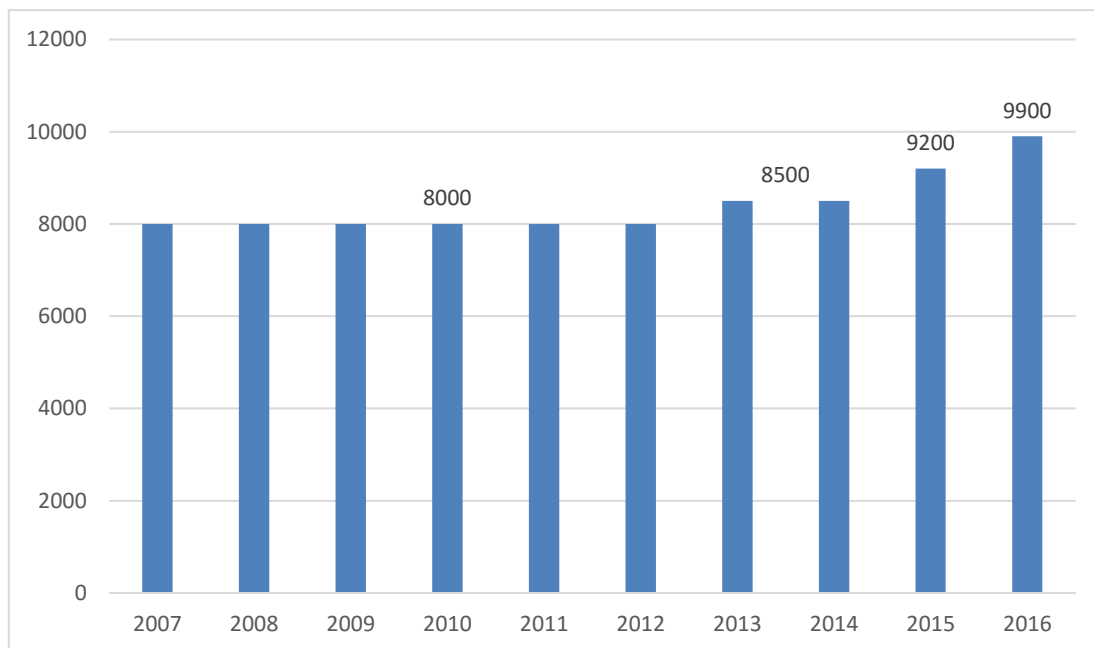


Zdroj: Česká strana sociálně demokratická (2015)

Na první pohled graf vzbuzuje dojem, že po stagnaci minimální mzdy v letech 2007–2012 došlo k výraznému nárůstu mezi lety 2013–2015, který pokračuje i v návrhu pro rok 2016. Nelze vyvrátit skutečnost, že Vláda České republiky ve sledovaném období opravdu minimální mzdu zvýšila. Problém pramení z toho, že osa Y začíná na hodnotě 7 700 Kč a jednotlivé změny v minimální mzdě tak na první pohled vykazují rapidní růst. V konečném důsledku pak graf vzbuzuje dojem, že je hodnota minimálního příjmu v letech 2007–2012 na velmi nízké úrovni, a naopak v roce 2016 opravdu vysoko.

V Grafu 1 jsou zobrazena stejná data, ale v korektním formátu. Osa Y je zobrazena v celém intervalu a jsou přidány vodící linky, které čtenáři umožní snadné porovnávání hodnot v jednotlivých letech.

Graf 1 – Vývoj minimální mzdy v ČR od roku 2007

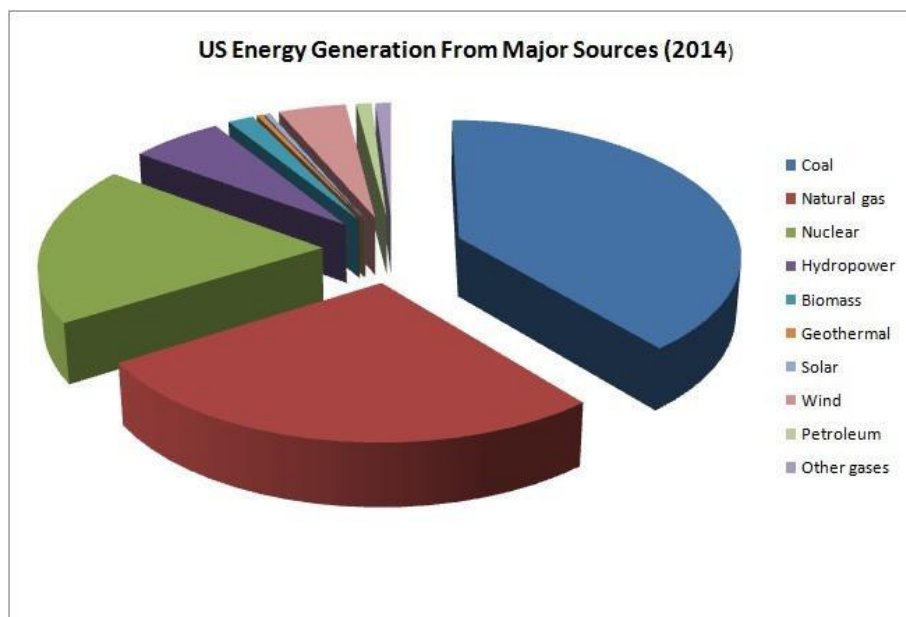


Zdroj: Česká strana sociálně demokratická (2015), vlastní zpracování

Podobným způsobem jako sloupcový graf může být k manipulaci zneužit také graf koláčový. Ten je nejčastěji používán pro vyjádření podílu určitých hodnot na daném celku. U tohoto grafu může být sdělení, které z něj má pramenit, upraveno prostřednictvím jeho prostorového otočení, oddělením jednotlivých částí grafu či případně opomenutím popisků jednotlivých hodnot.

Výše zmíněné tři prohřešky byly použity v grafu, který se nachází ve článku s názvem „*Electric Vehicles And The End Of The Oil Majors*“. Autor tohoto článku, který se objevil na serveru Seeking Alpha, pomocí koláčového grafu vyjadřuje podíl jednotlivých zdrojů na celkové americké produkci elektrické energie. Mezi chyby použité při konstrukci tohoto grafu (Obrázek 12) patří chybějící hodnoty podílů, mírně oddálené jednotlivé výseky a 3D efekt, který na první pohled zkresluje porovnání velikosti jednotlivých zdrojů elektrické energie. Kvůli velkému počtu skupin, které se v grafu nacházejí, není také přesně zřejmé, který popisek náleží k jaké výseči.

Obrázek 12 – Příklad zavádějícího formátu koláčového grafu

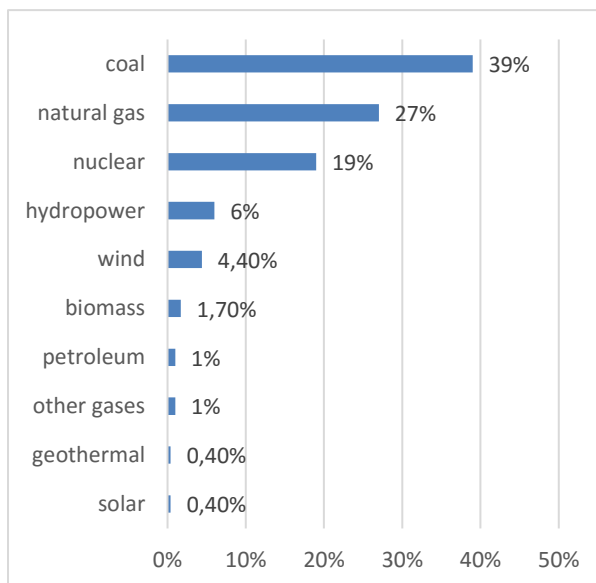


Zdroj: The Struggling Millennial (2016)

Z důvodu chybějících popisků a 3D efektu je na první pohled podíl uhlí a zemního plynu na výrobě energie v USA prakticky stejný. Ve skutečnosti se však liší o 12 procentních bodů. Pro zobrazení takovýchto dat by se nabízel spíše sloupcový graf či tabulka s jednotlivými hodnotami (Tabulka 9). Z takto prezentovaných dat je jasně zřejmý podíl jednotlivých částí na celku.

Tabulka 9 – Zdroje pro výrobu energie v USA v roce 2014

Zdroj	Podíl
Coal	39 %
Natural gas	27 %
Nuclear	19 %
Hydropower	6 %
Wind	4,4 %
Biomass	1,7 %
Petroleum	1 %
Other gases	1 %
Geothermal	0,4 %
Solar	0,4 %



Zdroj: The Struggling Millennial (2016)

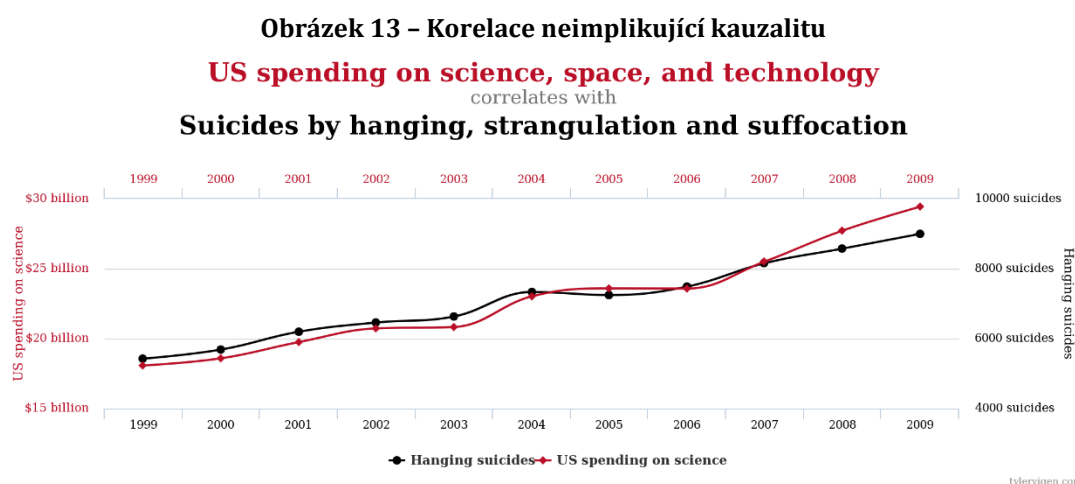
Předchozí příklady znázorňují, jak jednoduché je upravit graf tak, aby se jeho sdělení markantně lišilo od skutečnosti, kterou by měl reflektovat. Z tohoto důvodu je nezbytné mít na paměti, že jakýkoliv graf bez ohledu na to, zda se jedná o sloupcový, výsečový nebo spojnicový, může být upraven za účelem manipulace. Proto je vhodné si k jasné vypadajícím grafům zjistit také další podrobnosti o datech.

4.4 Klam narativity

Esejista Nassim Taleb v jedné z kapitol své knihy Černá labuť vysvětluje, že lidský mozek má sklony k vytváření konzistentních příběhů a dává do vzájemné souvislosti jevy, které spolu ve skutečnosti vůbec nesouvisejí. Pokud se dva jevy vyskytují zároveň, následují bezprostředně po sobě nebo jsou spolu ve vzájemné korelaci, není ještě možné usuzovat, že jeden jev je příčinou toho druhého. Taleb tento jev nazval klamem narativity, ale ve statistických člancích či knihách je zdůrazňováno, že korelace nemusí nutně implikovat kauzalitu.

Na první pohled srozumitelná myšlenka. Pokud se v grafu objeví dvě křivky, přičemž jedna téměř přesně kopíruje tu druhou, snadno se čtenář nechá zmást. V jeho mysli se objeví myšlenka, že oba jevy spolu mohou vzájemně souviset.

Vigen (2015) ve své knize Spurious Correlations zobrazuje několik příkladů, ve kterých vysoká korelace neznamena v žádném případě kauzalitu daných jevů. Na Obrázku 13 je zobrazen vývoj amerických výdajů na vědu, vesmírný průzkum a technologie a počtu vražd spáchaných oběšením, uškrcením nebo udušením.



Zdroj: Vigen (2015)

Ačkoliv je vývoj obou ukazatelů na první pohled téměř stejný, oba ukazatele spolu absolutně nesouvisejí. Stačí však vhodně zvolené měřítko a sledované období a rázem začne mozek hledat v daných jevech vzájemné souvislosti.

Jak dokazují příklady výše, úskalí manipulace za využití statistických metod nebo nástrojů se může objevit opravdu ve všech oblastech lidské činnosti a velmi snadno zmást neostražitěho čtenáře. Huff (1954) doporučuje položit při konzumaci jakéhokoliv obsahu 5 základních otázek, které pomohou posoudit, zda se jedná o věrohodná data, a to:

- Kdo to říká?
- Jak to ví?
- Co chybí?
- Nezměnil někdo předmět?
- Dává to smysl?

5 Dotazníkové šetření a analýza dat

Ústředním prvkem praktické části diplomové práce je dotazníkové šetření a zejména analýza dat nasbíraných prostřednictvím tohoto dotazníku. Cílem provedené analýzy dat je nalézt odpovědi na výzkumné otázky stanovené v úvodu diplomové práce. Struktura dotazníkového šetření je ve stejné formě, ve které byla nastíněna v metodologické části práce. První sekce dotazníku prověřuje, zda respondenti znají základní statistické pojmy. Tato znalost není zjišťována pouze prostřednictvím strohých definic, ale také na základě jednoduchých příkladů. Druhá sekce se zabývá schopností respondentů interpretovat zprávy a informace, které jsou ne vždy předkládány v řádné formě. Třetí část obsahuje otázky týkající se základních osobních charakteristik respondentů. Základní analýza jednotlivých proměnných bude prováděna prostřednictvím programu Microsoft Excel 2016. Pro analýzu závislostí je zvolen statistický program SAS Enterprise Guide 5.1.

Dotazníkové šetření nese stejný název jako je titulek pravděpodobně nejznámější knihy zabývající se možným zneužitím statistických metod při ovlivnění veřejného mínění. Tuto knihu napsal Darrell Huff a nazval ji výstižně „*Jak lhát se statistikou?*“. Dotazník byl dostupný k vyplnění prostřednictvím webových stránek Vyplňto.cz v období od 23. 04. 2017 do 14. 05. 2017. Za tuto dobu se podařilo získat celkem 219 respondentů. Průměrná doba vyplňování byla 7 minut a 27 sekund a návratnost dotazníku 54,2 %. Samotná distribuce dotazníku probíhala několika různými toky. Prvním a hlavním distribučním kanálem byla sociální síť Facebook. Díky tomu, že Vyplňto.cz disponuje vlastní aplikací pro platformy iOS a Android, bylo možné provádět terénní sběr dat také elektronicky, což velmi usnadnilo a zrychlilo přípravou fázi datového souboru.

V dotazníku bylo použito celkem 14 otázek, z toho 13 se týkalo všech respondentů a jedna zjišťovala obor, ve kterém pracující respondenti vykonávají výdělečnou činnost. Jednotlivé otázky byly buď uzavřené, nebo polouzavřené, neboť povaha otázek nevyžadovala použití otevřeného typu otázek. Kompletní dotazník v podobě, ve které byl předkládán respondentům, se nachází mezi přílohami na konci diplomové práce.

5.1 Výsledky dotazníkového šetření

Před samotným zjišťováním závislostí mezi jednotlivými prvky dotazníku je nutně daný soubor blíže charakterizovat. V této části praktické části proto jsou představeny jednotlivé otázky a k nim přiřazena absolutní a relativní četnost odpovědí.

Jednotlivé otázky byly rozděleny do tří sekcí, proto se bude i tato charakteristika datového souboru členit do tří částí.

5.1.1 Charakteristika respondentů

Dotazníkového šetření se zúčastnilo celkem 219 respondentů různého sociálního postavení z různých krajů České republiky. Otázky osobního charakteru jsou na doporučení autorů příslušné literatury situované v poslední části dotazníku, zde jsou však pro lepší přehlednost analyzovány přednostně. V této sekci jich je celkem 6, přičemž na jednu otázku odpověděli pouze pracující respondenti.

Otázka číslo 1: „Jaké je Vaše pohlaví?“

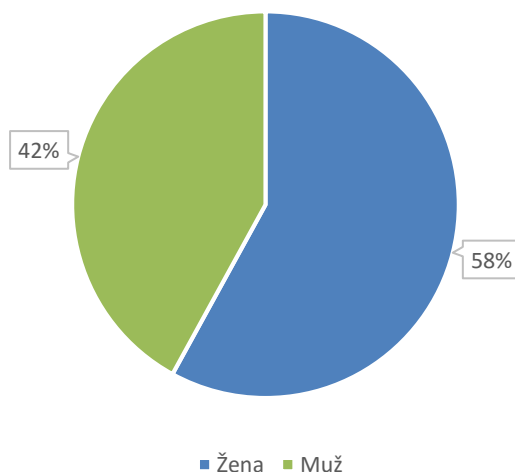
Tabulka 10 – Pohlaví respondentů

Pohlaví	Absolutní četnost	Relativní četnost
Žena	127	57,99 %
Muž	92	42,01 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Z Tabulky 10 vyplývá, že se mezi respondenty nachází více žen. Z absolutního počtu 219 dotázaných je 127 žen, což odpovídá relativní četnosti 57,99 %. Pro lepší přehlednost jsou data zobrazena také v koláčovém grafu (viz Graf 2).

Graf 2 – Struktura respondentů podle pohlaví



Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Otázka číslo 2: „Do jaké věkové kategorie se řadíte?“

Otázka číslo 2 se týká věku respondenta, kdy měl každý dotazovaný na výběr celkem ze čtyřech kategorií (viz Tabulka 11). Nutno podotknout, že spodní hranice věku respondenta byla stanovena na 15 let.

Tabulka 11 – Věkové složení respondentů

Věk	Absolutní četnost	Relativní četnost
Do 30 let	153	69,86 %
31–45 let	39	17,81 %
46–60 let	23	10,50 %
61 a více let	4	1,83 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Z Tabulky 11 je zřejmé, že nejpočetněji je zastoupena nejmladší skupina dotázaných, kdy je téměř 70 % respondentů mladších 31 let. Toto vychýlení vzorku je jednou z překážek toho, aby mohly být závěry z analýzy zobecněny na celou populaci. Výsledek je způsoben volbou distribučního kanálu, protože hlavním prostředkem byl internet, k němuž mají blíže mladí lidé. Toto vychýlení bohužel příliš nezmírnilo ani terénní šetření, které bylo v rámci sběru dat také provedeno. Nejméně zastoupenou byla věková skupina nad 61 let, která tvořila pouze 1,83 % z celkového počtu účastníků dotazníku.

Otázka číslo 3: „Jaké je Vaše nejvyšší dosažené vzdělání?“

Tabulka 12 – Nejvyšší dosažené vzdělání

Vzdělání	Absolutní četnost	Relativní četnost
Základní	10	4,57 %
Středoškolské bez maturity	7	3,20 %
Středoškolské s maturitou	109	49,77 %
Vyšší odborné	9	4,11 %
Vysokoškolské	84	38,36 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Vzdělání respondentů je rozděleno do 5 kategorií, přičemž nejpočetněji jsou zastoupeni lidé, kteří zakončili střední vzdělání s maturitou. Ze všech respondentů má toto vzdělání téměř každý druhý (viz Tabulka 12). Druhou hojně zastoupenou skupinou

jsou vysokoškolsky vzdělaní lidé. Ti v celkovém souboru tvoří 38,36 %. Ostatní skupiny jsou zastoupeny velmi zřídka a jejich relativní četnost se pohybuje mezi 3,20 % a 4,57 %.

Otázka číslo 4: „Jaký je Váš status?“

V rámci statusu volili respondenti z celkem 6 kategorií – zaměstnaný, nezaměstnaný, student, OSVČ, důchodce a jiný.

Tabulka 13 – Status respondentů

Status	Absolutní četnost	Relativní četnost
Student	96	43,84 %
OSVČ	11	5,02 %
Zaměstnaný	98	44,75 %
Nezaměstnaný	3	1,37 %
Důchodce	7	3,20 %
Jiný	4	1,83 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Nejčastěji se dotazníkového šetření účastnili zaměstnanci, kterých je 44,75 % a 98 absolutně. Pouze o 2 respondenty méně se nachází ve skupině studentů. Další 4 kategorie byly zastoupeny velmi slabě a pro účely dalších analýz je nutné jejich sloučení do jedné společné kategorie.

Otázka číslo 5: „V jakém oboru jste zaměstnaný?“

Otázka číslo 5 doplňuje otázku předchozí. Tato se však netýká všech respondentů, kteří se účastnili šetření, nýbrž pouze těch, kteří se nacházejí v zaměstnaneckém poměru nebo podnikají jako osoby samostatně výdělečně činné. Nutno podotknout, že se jedná o jedinou polouzavřenou otázku, tudíž respondenti měli možnost napsat libovolný obor či odvětví. Kvůli tomu jsou některé odpovědi zastoupeny pouze jednou a byla nutná jejich agregace (viz Tabulky 14). Došlo tak například ke sloučení odpovědí „učitelka“ a „pedagogika“ do souhrnné skupiny nazvané „školství“.

Samostatnou výdělečnou činnost vykonává 11 respondentů a v zaměstnaneckém poměru pracuje 98 respondentů. Otázka č. 5 se tedy týkala celkem 109 dotázaných, přičemž nejčastěji je zastoupen obor administrativa, v jehož rámci vykonává svoji činnost 24 ze 109 dotázaných.

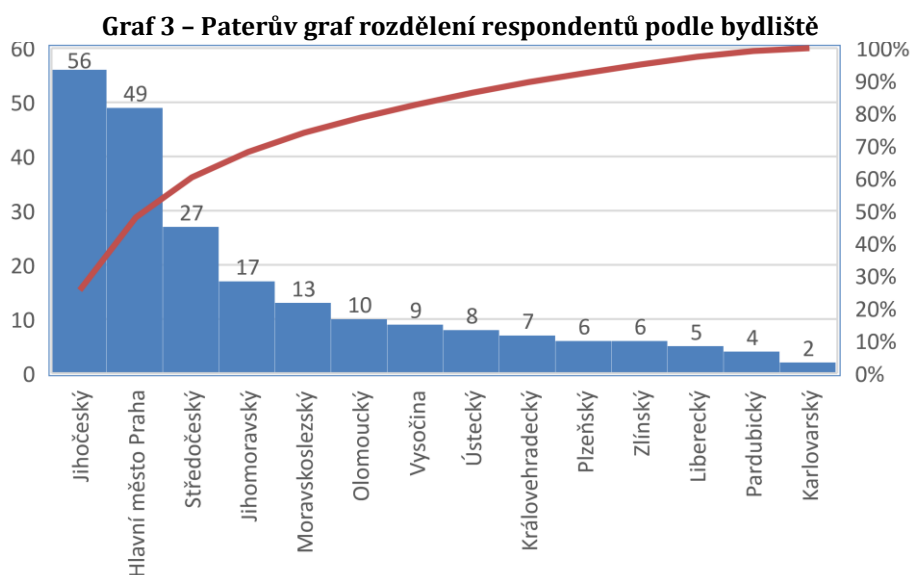
Tabulka 14 – Respondenti podle oboru

Obor	Absolutní četnost	Relativní četnost
Administrativa	24	22,02 %
Obchod	7	6,42 %
IT	12	11,01 %
Veřejná správa	16	14,68 %
Služby	17	15,60 %
Průmysl	18	16,51 %
Školství	7	6,42 %
Ostatní	8	7,34 %
Celkem	109	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Otázka číslo 6: „V jakém kraji bydlíte?“

Poslední otázka této sekce se týká bydliště respondentů, kdy je na výběr jeden ze 14 krajů České republiky. Jednotlivá zastoupení jsou velmi proměnlivá. Více než polovina respondentů pochází z Prahy, Jihočeského nebo Středočeského kraje. To potvrzuje skutečnost, že není možné zobecnit výsledky dotazníku na celou populaci Čechů, jelikož vzorek respondentů nelze považovat za reprezentativní vzorek obyvatel České republiky. Jednotlivé četnosti jsou zobrazeny v Tabulce 15 a pro lepší přehlednost rovněž v Paterově grafu (viz Graf 3), ve kterém je zřejmý také kumulativní podíl.



Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 15 – Bydliště respondentů

Bydliště	Absolutní četnost	Relativní četnost
Hlavní město Praha	49	22,37 %
Středočeský	27	12,33 %
Jihočeský	56	25,57 %
Plzeňský	6	2,74 %
Karlovarský	2	0,91 %
Ústecký	8	3,65 %
Liberecký	5	2,28 %
Královehradecký	7	3,20 %
Pardubický	4	1,83 %
Vysočina	9	4,11 %
Jihomoravský	17	7,76 %
Olomoucký	10	4,57 %
Zlínský	6	2,74 %
Moravskoslezský	13	5,94 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

5.1.2 Analýza odpovědí na základní statistické pojmy

Druhou část dotazníku tvoří otázky, na jejichž základě je možné hodnotit znalost respondentů v oblasti statistických pojmů. Ve čtyřech samostatných otázkách museli respondenti prokázat znalost pojmů průměr, medián, modus a rozptyl. U pojmů charakterizujících míru polohy hodnot se jednalo o praktickou otázku v podobě jednoduchého příkladu. Typ průměru byl v rámci dotazníku blíže specifikován a jednalo se o průměr aritmetický.

Otázky týkající se charakteristik míry polohy vycházely z totožného příkladu, kdy bylo stanoveno 5 hodnot (3;1;3;3;2) a respondent měl postupně určit průměr, modus a medián. Ze zadaných hodnot nabývá průměr hodnoty 2,4, modus 3 a medián také 3.

Otázka číslo 7: „Kolik je průměr?“

Tabulka 16 – „Kolik je průměr?“

Hodnota	Absolutní četnost	Relativní četnost
3	3	1,37 %
2,4	209	95,43 %
12	1	0,46 %
2	2	0,91 %
Nevím	4	1,83 %
1	0	0,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Otázka týkající se aritmetického průměru činila respondentům dle předpokladů nejménší problém. Z celkového počtu 219 dotázaných dokáže tento ukazatel určit 95,43 % a pouhých 10 respondentů tento ukazatel určit nedokáže. Vysoká úspěšnost odpovědí plyne z častého používání této statistické veličiny a její jednoduché a notoricky známé konstrukce. Z toho plyne snadná zneužitelnost, která byla zmíněna v části práce zabývající se zneužitím statistických pojmů.

Otázka číslo 8: „Kolik je modus?“

Pojem modus je pro respondenty mnohem obtížnější než průměr. Tento ukazatel dovede správně stanovit 74,43 % respondentů, 18,26 % dotázaných přiznává, že „neví“ a 7,31 % tipuje nesprávný výsledek (viz Tabulka 17).

Tabulka 17 – „Kolik je modus?“

Hodnota	Absolutní četnost	Relativní četnost
3	163	74,43 %
2,4	3	1,37 %
12	8	3,65 %
1	1	0,46 %
2	4	1,83 %
Nevím	40	18,26 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Otázka číslo 9: „Kolik je medián?“

Ještě hůře je na tom znalost týkající se mediánu. Tento relativně známý a běžně používaný pojem, který má spousta osob spojena s vyjádřením mediánové hodnoty

příjmu, je schopno správně určit pouze 63,93 % dotázaných. Možnost nevím je zastoupena 12,33 % a 23,75 % respondentů odpovídá špatně.

Tabulka 18 – „Kolik je medián?“

Hodnota	Absolutní četnost	Relativní četnost
3	140	63,93 %
2,4	7	3,20 %
12	0	0,00 %
1	3	1,37 %
2	42	19,18 %
Nevím	27	12,33 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Velmi zajímavé je srovnání výsledků mezi módem a mediánem, kdy v případě módu dala větší část dotázaných přednost odpovědi „nevím“ před určením špatné hodnoty. Naopak tomu bylo u mediánu, zde větší část respondentů určila špatnou hodnotu než odpověď „nevím“. Tyto výsledky naznačují, že pojem medián je mezi lidmi známější, ale jeho výklad je často chybný.

Otázka číslo 10: „Co je to rozptyl?“

Otázka 10 již není praktického charakteru, ale respondentům byly předloženy různé definice, ze kterých bylo nutné zvolit tu, která definovala rozptyl. Úspěšnost této otázky je vysoká, což dokládají také hodnoty v Tabulce 19.

Tabulka 19 - „Co je to rozptyl?“

Odpověď	Absolutní četnost	Relativní četnost
Hodnota, která se v daném souboru vyskytuje nejčastěji.	2	0,91 %
Hodnota, která vyjadřuje, jak moc hodnoty souboru kolísají kolem střední hodnoty.	203	92,69 %
Nevím.	10	4,57 %
Součet všech hodnot souboru vydělený jejich celkovým počtem.	4	1,83 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Vysoká znalost vycházející z dat vyplývá z typu otázky, kdy je odpověď prostřednictvím definice pro respondenty snazší. Ověření znalosti tohoto pojmu na základě výpočtu byla u této otázky z důvodu obtížnější konstrukce vyloučena. Příliš složitě

otázky by mohly vést ke ztrátě zájmu a ukončení dotazníku před jeho kompletním vyplněním.

Na základě jednotlivých odpovědí, které jsou součástí této sekce dotazníku, je možné určit, kolik pojmů dokázal daný jedinec určit správně a kolik špatně. Pro analýzu závislosti znalosti pojmů statistiky na určitých faktorech či analýze závislosti schopnosti interpretace informací na znalosti pojmů statistiky je nutné shrnout tyto čtyři otázky do jednoho ukazatele. Jednotliví respondenti jsou rozděleni podle počtu správných odpovědí do 4 skupin (viz Tabulka 20).

Tabulka 20 – Znalost základních statistických pojmů

Počet správných odpovědí	Znalost statistických pojmů	Absolutní četnost	Relativní četnost
0–1	Špatná	12	5,48 %
2	Částečná	35	15,98 %
3	Dobrá	52	23,74 %
4	Výborná	120	54,79 %
Celkem	-	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Z tabulky četností vyplývá, že podle stanovených parametrů statistickým pojmům výborně rozumí 54,79 % respondentů. Do této skupiny jsou zařazeni respondenti, kteří dokáží správně určit všechny 4 pojmy. Naopak 12 dotázaných nedokázalo odpovědět ani na jedinou otázku.

5.1.3 Analýza schopnosti interpretace informačních sdělení

V třetí části dotazníkového šetření byly respondentům předkládány informace ve formě novinových titulků a sloupcových či koláčových grafů. Úkolem dotazovaného bylo z takto podaných informací vyvodit správné závěry, které se nacházely jako jedna z možností mezi ostatními odpověďmi. Stejně jako předchozí část i tato obsahuje 4 otázky. Jedna je zaměřena na schopnost interpretace informací vyjádřených prostřednictvím procent. Ve druhé se respondenti snaží odhalit chybné závěry vyvozené z nereprezentativního vzorku populace. Otázka třetí a čtvrtá prověřuje schopnost interpretace informací pocházející z upraveného koláčového a sloupcového grafu.

Otázka číslo 11: „Odráží novinový titulek skutečnost?“

Na první pohled snadná otázka zaměřená na počítání s procenty se stala pro respondenty velmi obtížnou. Z celkového počtu 219 dotázaných je schopno na tuto otázku

odpovědět správně 96 z nich, což je v relativním vyjádření pouhých 43,84 %. Výsledky jsou zaneseny v Tabulce 21 a úplné znění otázky se nachází v Příloze č. 2.

Tabulka 21 - „Odráží novinový titulek skutečnost?“

Odpověď	Absolutní četnost	Relativní četnost
Novinový titulek odráží skutečnost. Cena nemovitostí se opravdu nezměnila.	38	17,35 %
Novinový titulek neodráží skutečnost. Ceny nemovitostí v roce 2017 jsou nižší než v roce 2015.	96	43,84 %
Novinový titulek neodráží skutečnost. Ceny nemovitostí v roce 2017 jsou vyšší než v roce 2015.	73	33,33 %
Nevím.	12	5,48 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Otázka číslo 12: „Je na základě provedeného průzkumu možno spolehlivě určit míru životní spokojenosti občanů ČR?“

Následující úkol prověřoval schopnost dotázaných odhalit chybně zobecněné závěry stanovené na základě vychýleného vzorku občanů. Podobně koncipovaná šetření se v novinách či jiných informačních médiích objevují velmi často. Každý příjemce zprávy, která zobecňuje poznatky na celou populaci či část populace, se musí ptát, na základě jakého výběrového souboru jsou tyto závěry vyvozeny. Respondenti účastníci se dotazníkového šetření tento problém odhalili poměrně dobře. Správnou odpověď zvolilo 85,39 % z nich (viz Tabulka 22).

Tabulka 22 - „Je na základě provedeného průzkumu možno spolehlivě určit míru životní spokojenosti občanů ČR?“

Odpověď	Absolutní četnost	Relativní četnost
Na základě provedeného průzkumu není možné stanovit skutečnou spokojenost obyvatel České republiky se svým životem.	187	85,39 %
Provedený průzkum lze považovat za věrohodný. Se svým životem může být spokojeno 85 % obyvatel České republiky.	8	3,65 %
Průzkum je smyšlený, žádný ze jmenovaných zařízení by nedovolilo pořádání průzkumu.	16	7,31 %
Nevím.	8	3,65 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Otázka číslo 13: „Co plyne z koláčového grafu?“

Otázka zaměřená na schopnost interpretovat data obsažená v záměrně upraveném koláčovém grafu je jedinou otázkou dotazníku, která obsahovala dvě správné odpovědi. Koláčový graf, podobně jako některé koláčové grafy prezentované například televizní stanicí FOX News, je záměrně upraven. Úprava spočívá v tom, že jedna z výsečí koláčového grafu svou velikostí neodpovídá skutečnému podílu. Konkrétně podíl 37 % je v koláčovém grafu reprezentován výsečí o velikosti přibližně 55 %.

Tabulka 23 – Schopnost interpretovat koláčový graf

Odpověď	Absolutní četnost	Relativní četnost
Správně	123	56,16 %
Špatně	90	41,10 %
Nevím	6	2,74 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Pole „*správně*“ v Tabulce 23 reprezentuje ty respondenty, kteří z daných odpovědí zvolili obě správné odpovědi. V případě, že některý zvolil pouze jednu, byl zařazen do kategorie špatně odpovídajících.

Otázka číslo 14: „Co vyplývá ze sloupcového grafu?“

Poslední otázka dotazníků prověřuje schopnost interpretace informací ze sloupcového grafu, u kterého došlo k oříznutí osy y. Interval hodnot osy y nezačíná na hodnotě 0, nýbrž na hodnotě 94 milionů. Tím je vytvořen dojem, že dochází ke značnému růstu dané hodnoty, který je ve skutečnosti v řádu jednotek procent (viz Příloha č. 1). Interpretaci sloupcového grafu považuje mnoho autorů za snazší než interpretaci koláčového grafu. Toto tvrzení potvrzuje také provedený průzkum, ve kterém tento graf dokázalo správně interpretovat 178 respondentů (viz Tabulka 24). Nutno podotknout, že stejně jako u předchozí otázky bylo možné zvolit více odpovědí, zde však správně byla pouze jedna z nich.

Tabulka 24 – Schopnost interpretovat sloupcový graf

Odpověď	Absolutní četnost	Relativní četnost
Správně	178	81,28 %
Špatně	29	13,24 %
Nevím	12	5,48 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Stejně jako v části zaměřené na znalost statistických pojmů, je nutné i tuto část pro účely následujících analýz shrnout do jednoho ukazatele. Parametr počtu správných odpovědí pro přiřazení respondentů do konkrétních skupin je stejný jako v předchozím případě.

Tabulka 25 – Schopnost interpretace informačních sdělení

Počet správných odpovědí	Schopnost interpretace	Absolutní četnost	Relativní četnost
0–1	Špatná	33	15,07 %
2	Částečná	59	26,94 %
3	Dobrá	70	31,96 %
4	Výborná	57	26,03 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Z Tabulky 25 vyplývá, že celkem 127 dotázaných dokáže dobře nebo výborně interpretovat informace, se kterými se v běžném životě mohou setkat.

5.2 Analýza závislostí

Po provedené přípravě dat, která proběhla v rámci předchozí části diplomové práce, je možné provést analýzu závislostí. Tato analýza si klade za cíl zodpovědět výzkumné otázky stanovené v úvodu diplomové práce. Celkem budou provedeny dvě analýzy. První se bude zabývat závislostí znalosti základních statistických pojmů na jednotlivých osobních charakteristikách respondentů. Cílem druhé analýzy je zjištění, zda závisí schopnost interpretovat zprávy na znalosti statistických pojmů.

5.2.1 Závislost znalosti statistických pojmů na daných proměnných

Pro zjištění závislosti znalosti statistických pojmů na osobních charakteristikách respondentů bude Chí-kvadrát test nezávislosti, kdy bude postupně prověřována závislost mezi jednotlivými proměnnými. Aby byl tento záměr proveditelný, byla závisle proměnná znalost statistických pojmů v předchozí části práce rozdělena do 4 kategorií na základě počtu správných odpovědí (viz Tabulka 26). Veškeré testy budou prováděny na 95% hladině spolehlivosti. Pro zajištění snazší orientace v textu se výsledky jednotlivých testů nacházejí v Příloze č. 1.

Tabulka 26 – Znalost statistických pojmů

Počet správných odpovědí	Zástupný symbol	Znalost statistických pojmů
0-1	0	Špatná
2	1	Částečná
3	2	Dobrá
4	3	Výborná

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Nezávisle proměnných veličin, které budou postupně prověřovány, je celkem 5, a to:

- pohlaví,
- věk,
- bydliště,
- dosažené vzdělání a
- status.

Závislost znalosti statistických pojmů na pohlaví

- H_0 : Znalost statistických pojmů nezávisí na pohlaví.
- H_1 : Znalost statistických pojmů závisí na pohlaví.

P-hodnota provedeného testu dosahuje 67,22 % a je tedy vysoce nad 5% hladinou významnosti. Na 5% hladině významnosti tedy není prokázána závislost znalosti statistických pojmů na pohlaví. Tento výsledek se dal předpokládat, jelikož pohlaví nemá vliv na znalosti a inteligenci.

Závislost znalosti statistických pojmů na věku

- H_0 : Znalost statistických pojmů nezávisí na věku.
- H_1 : Znalost statistických pojmů závisí na věku.

První pokus provedení testu není úspěšný, jelikož nebyla splněna podmínka, že minimálně 80 % buněk kontingenční tabulky musí dosahovat minimální četnosti 5. Aby bylo možné test provést, došlo k redukci obou proměnných. Proměnná znalost statistických pojmů je zredukována do dvou kategorií (viz Tabulka 27).

Proměnná věk nabývala před úpravou celkem 4 kategorií, přičemž kategorie zahrnující respondenty nad 61 let obsahovala pouze 4 dotázané. Z toho důvodu byla tato skupina sloučena se skupinou občanů ve věku 46-60 let, výsledné četnosti jsou uvedeny v Tabulce 28.

Tabulka 27 – Znalost statistických pojmů po úpravě

Počet správných odpovědí	Znalost	Absolutní četnost	Relativní četnost
0-2	ne	47	21,46 %
3-4	ano	172	78,54 %
Celkem	-	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 28 – Četnost věkových kategorií po úpravě

Věk	Absolutní četnost	Relativní četnost
Do 30 let	153	69,86 %
31-45 let	39	17,81 %
46 a více let	27	12,33 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Uskutečněné úpravy již umožňují provedení daného testu. Výsledná p-hodnota (2,55 %) je tentokrát nižší než hladina významnosti. Znalost statistických pojmů je tedy závislá na věku respondenta, přičemž s rostoucím věkem se tato znalost snižuje (viz Příloha č. 1).

Závislost znalosti statistických pojmů na bydlišti

- H_0 : Znalost statistických pojmů nezávisí na bydlišti respondenta.
- H_1 : Znalost statistických pojmů závisí na bydlišti respondenta.

Kvůli malému zastoupení respondentů z některých krajů, byla ještě testováním tato proměnná sloučena do dvou skupin na základě historických území České republiky (viz Tabulka 29). První skupina je tvořena kraji Čech a ve druhé skupině se nacházejí kraje Moravy a Slezska. Jelikož hranice jednotlivých území nekopíruje hranice krajů, je Vysočina zařazena do Moravy, Pardubický a Jihočeský kraj do Čech. Slezsko je reprezentováno pouze Moravskoslezským krajem.

Tabulka 29 – Bydliště respondentů dle historických území ČR

Bydliště	Absolutní četnost	Relativní četnost
Čechy	164	74,89 %
Morava a Slezsko	55	25,11 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Závisle proměnná je v rámci tohoto testu použita bez úpravy a nabývá tedy 4 hodnot podle úrovně znalosti statistických pojmů.

Výsledná p-hodnota stanovená pomocí Chí-kvadrátu testu nezávislosti nabývá hodnoty 24,23 %. Na jejím základě neprokazují alternativní hypotézu, že znalost statistických pojmů závisí na místě bydliště respondentů. Tento výsledek je také očekávaný, jelikož mezi jednotlivými oblastmi není výrazná geografická nebo kulturní vzdálenost, navíc se jedná o oblasti v rámci jedné republiky.

Závislost znalosti statistických pojmů na dosaženém vzdělání

- H_0 : Znalost statistických pojmů nezávisí na vzdělání respondenta.
- H_1 : Znalost statistických pojmů závisí na vzdělání respondenta.

Za účelem splnění četností alespoň u 80 % buněk kontingenční tabulky došlo ke sloučení vyššího odborného vzdělání se středoškolským vzděláním s maturitou, ke kterému mají tyto obory blíže než k univerzitám (viz Tabulka 30). Dále byly sloučeny respondenti se základním a středoškolským vzděláním bez maturity.

Tabulka 30 – Stupeň dosaženého vzdělání po úpravě

Vzdělání	Absolutní četnost	Relativní četnost
Základní + středoškolské bez maturity	17	7,76 %
Středoškolské s maturitou	118	53,88 %
Vysokoškolské	84	38,36 %
Celkový součet	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Proměnná reprezentující znalost statistických pojmů je použita opět v agregované formě, kdy nabývá pouze dvou hodnot (viz Tabulka 27).

Výsledná p-hodnota (0,04 %) prokazuje na 5% hladině významnosti platnost alternativní hypotézy. Mezi znalostí statistických pojmů a vzděláním panuje přímá úměra, kdy se se zvyšující úrovní vzdělání zvyšuje i znalost respondentů v oblasti statistiky.

Závislost znalosti statistických pojmů na statusu respondenta

- H_0 : Znalost statistických pojmů nezávisí na statusu.
- H_1 : Znalost statistických pojmů závisí na statusu.

Poslední nezávislou proměnnou, která bude v souvislosti se znalostmi statistických pojmů testována je status respondentů. U proměnné status došlo ke sloučení hodnot nezaměstnaný, důchodce a jiný do jedné kategorie s názvem ostatní. Zároveň také

skupiny zaměstnaný a OSVČ do nové kategorie s názvem osoby výdělečně činné (viz Tabulka 31).

Tabulka 31 - Četnost statusu po úpravě

Status	Absolutní četnost	Relativní četnost
Student	96	43,84 %
Výdělečně činný	109	49,77 %
Ostatní	14	6,39 %
Celkem	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Proměnná zastupující znalost statistických pojmů je použita opět v agregované formě a nabývá pouze dvou hodnot (viz Tabulka 27, s. 64).

Výsledná hodnota Chí-kvadrátu testu nezávislosti je menší než 0,01 %. Na 95% hladině spolehlivosti je možné prokázat platnost alternativní hypotézy H_1 . Znalost statistických pojmů tedy závisí na statusu respondentů dotazníkového šetření. Nejúspěšnější byla skupina studentů, ze kterých prokázalo znalost statistických pojmů téměř 90 %. Naopak velmi špatné výsledky vykazuje skupina ostatní, která zahrnuje důchodce, nezaměstnané a ostatní respondenty. Správně nedokázala odpovědět ani polovina respondentů této skupiny.

5.2.2 Závislost schopnosti interpretovat zprávy na znalosti statistických pojmů

V úvodu diplomové práce byla stanovena výzkumná otázka, zda závisí schopnost správné interpretace informací na znalostech statistických pojmů. Odpověď na tuto otázku bude vycházet z výsledků dotazníkového šetření, kdy bude na základě Chí-kvadrátu testu nezávislosti posouzena závislost schopnosti interpretovat předložené zprávy na znalosti základních statistických pojmů.

Před samotným testováním je nutné stanovit výzkumné hypotézy:

- H_0 : Schopnost interpretace nezávisí na znalosti statistických pojmů.
- H_1 : Schopnost interpretace závisí na znalosti statistických pojmů.

Při použití neupravených dat nebyla splněna podmínka, že minimálně 80 % buněk kontingenční tabulky má minimální nebo očekávanou četnost 5 a více (viz Příloha č. 1). Za účelem splnění tohoto předpokladu došlo ke sloučení dat do čtyřpolní tabulky. Obě proměnné nyní nabývají pouhých 2 hodnot opět podle počtu správných odpovědí na dané otázky dotazníku (viz Tabulka 30 a 31).

Tabulka 32 – Znalost statistických pojmů po úpravě

Počet správných odpovědí	Znalost	Absolutní četnost	Relativní četnost
0-2	ne	47	21,46 %
3-4	ano	172	78,54 %
Celkem	-	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 33 – Schopnost interpretace informací po úpravě

Počet správných odpovědí	Schopnost	Absolutní četnost	Relativní četnost
0-2	ne	92	42,01 %
3-4	ano	127	57,99 %
Celkem	-	219	100,00 %

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Po úpravě již jednotlivé četnosti splňují předpoklady pro použití tohoto druhu testu. P-hodnota vyšla $<0,00,1$ na základě čehož je možné na 5 % hladině významnosti prokázat, že schopnost správné interpretace informací prezentovaných pomocí různých prvků statistiky závisí na znalosti základních statistických pojmů.

Stejný výsledek potvrzuje také korelační analýza postavená na neupravených datech, kdy jednotlivé proměnné nabývají 4 kategorií. Síla korelačního koeficientu nabývá hodnoty 0,375, což podle de Vause (2002) značí střední až podstatnou souvislost mezi proměnnými. Z tohoto korelačního koeficientu vyplývá, že se zvyšujícím se počtem statistických pojmů, které dokáže respondent určit, stoupá také schopnost interpretovat informace obsažené ve zprávách.

5.3 Shrnutí výsledků dotazníkového šetření

Bohužel hned z úvodních analýz statistického souboru vyplývá značné vychýlení vzorku respondentů. V rámci populace České republiky není možné tento vzorek považovat za reprezentativní hned z několika důvodů. Zásadními proměnnými, které znemožňují zobecnění závěru na celou populaci Čechů, jsou však proměnné věková kategorie a status respondentů. Zvolení internetu jako hlavního distribučního kanálu dotazníku vedlo k tomu, že se dotazování zúčastnili hlavně občané ve věku do 30 let. Tuto skutečnost se bohužel nepodařilo zmírnit ani terénním sběrem dat pomocí mobilní aplikace. Věková skupina 61 a více let byla proto zastoupena pouze 4 respondenty. Druhou proměnnou, která je výrazně postižena touto chybou, je proměnná status respondenta, zejména skupina respondentů se statusem důchodce.

V České republice žije k září 2016 celkem 2 880 677 občanů pobírajících alespoň jeden z různých druhů důchodů, které vyplácí Česká správa sociálního zabezpečení. Při porovnání tohoto počtu s celkovým počtem obyvatel ČR, který činí podle Českého statistického úřadu 10 553 800 obyvatel, jejich podíl činí 27,30 % (Český statistický úřad, 2016). Je pravděpodobné, že určitý počet obyvatel pobírajících důchod také pracuje. Například v roce 2011 to bylo 248 900 osob (Český statistický úřad, 2012).

Ve statistickém souboru se nachází pouze 3,20 % důchodců, což je v porovnání s výše uvedeným podílem velmi odlišné. Důvodem tohoto vychýlení je opět volba internetu, jako hlavního distribučního kanálu, ke kterému se dostane menší poměr důchodců, než se nachází v populaci.

Vychýlení vykazuje také proměnná pohlaví a bydliště. V celkovém souboru se nachází více žen (57,99 %) než mužů (42,01 %). Z celkového množství 219 respondentů jich více než polovina žije v Jihočeském a Středočeském kraji nebo v hlavním městě České republiky. Tyto dvě proměnné byly však na základě testu nezávislosti vyhodnoceny jako faktory, které neovlivňují znalost statistických pojmů. Proto jejich vychýlení nemá takovou důležitost jako vychýlení předchozích dvou proměnných.

Ačkoliv se jedná o nereprezentativní soubor respondentů, je možné přesto provést statistické testy, které budou pro diplomovou práci přínosné. Jedním z cílů diplomové práce není zobecnit výsledky na celou populaci, ale posoudit výsledky v rámci skupiny dotázaných.

Před samotným testováním jednotlivých závislostí uvnitř souboru bylo nutné některé proměnné agregovat tak, aby splňovaly minimální četnosti v kontingenční tabulce. Na základě Chí-kvadrátu testu nezávislosti bylo zjištěno, že znalost statistických pojmů závisí na věku, vzdělání a statusu respondentů. Naopak u proměnných pohlaví a místo bydliště nebyla tato závislost prokázána.

Počet kategorií proměnné věk byl snížen na 3, kdy došlo ke sloučení nejstarších dvou skupin do jedné. Ze sestavené kontingenční tabulky a provedení Chí-kvadrátu testu nezávislosti plyne, že u vyšších věkových kategorií je znalost statistických pojmů nižší. Tento výsledek je způsoben z části tím, že v nejmladší kategorii se nacházel největší podíl studentů, kteří se s těmito pojmy setkávají nebo se v nedávné době setkali.

Proměnná vzdělání byla zredukována opět do tří skupin. Zde došlo ke sloučení respondentů s vyšším odborným vzděláním s těmi se středoškolským vzděláním s maturitou. Dále byli sloučeni respondenti se základním a středoškolským vzděláním bez maturity. Z výsledků testů vyplývá, že respondenti s vyšším dosaženým

vzděláním dosahují lepších výsledků v otázkách zaměřených na znalost statistických pojmů.

Třetí proměnná mající vliv na znalosti statistických pojmů je status respondentů. Tento faktor byl sloučen opět do třech kategorií. Z výsledků plyne, že téměř 90 % studentů odpovědělo na otázky směřující na jejich znalost statistických pojmů bezchybně, případně udělali pouze jednu chybu. Naopak velmi špatné výsledky vykazuje skupina ostatní, která zahrnuje důchodce, nezaměstnané a ostatní respondenty. Více než polovina (57,14 %) respondentů této skupiny neprokázala znalost statistických pojmů.

Hlavním testem, který byl v rámci práce proveden, bylo zjištění závislosti mezi schopností interpretovat informace obsažené ve zprávách, které mohou být určitým způsobem matoucí a znalostí statistických pojmů. Za účelem provedení testu nezávislosti došlo k úpravě dat obou skupin. Po této modifikaci nabývala každá z proměnných pouze dvou hodnot (schopen/neschopen a zná/nezná). Výsledný test prokázal závislost obou proměnných. Respondenti, kteří prokázali znalost statistických pojmů, jsou tedy lépe připraveni odolávat nástrahám informačních manipulátorů.

Závěr

Diplomová práce se zabývala možnostmi využití a případného zneužití statistiky k tvorbě informací a následného ovlivňování veřejného mínění. Aktuálnost tohoto tématu je způsobena novými možnostmi sběru, analýzy a zejména prezentace dat. Vývoj těchto prvků statistiky pramení z velmi dynamického rozvoje informačních technologií.

Teoretická část diplomové práce byla rozdělena na dva celky. První se zabýval pojmem veřejné mínění. Zde došlo k jeho vymezení, jak ze současného, tak i historického hlediska. Dále zde byl nastíněn proces tvorby veřejného mínění a způsoby, kterými se provádí jeho průzkumy.

Druhá kapitola teoretické části práce se již věnuje samotné statistice. Na počátku je opět stručně nastíněna definice a vývoj tohoto pojmu, načež navazuje představení jednotlivých způsobů analýz statistických dat. Jedná se zejména o představení možností analýzy jednotlivých proměnných a testování hypotéz. Jednotlivé přístupy byly následně využity v praktické části práce k rozboru a analyzování proměnných v rámci dotazníkového šetření.

Praktická část začíná stručnou metodikou, která je zaměřena zejména na problematiku dotazování a dotazníkového šetření. Následně jsou nastíněny postupy, ve kterých je statistika nebo její prvky použity k předávání informací a tvorbě nebo ovlivňování veřejného mínění.

Prvním ze skupiny možných chyb při tvorbě informací je problém nereprezentativního vzorku, kdy se původce informací snaží zobecnit závěry na základě dat od vychýleného vzorku respondentů. Jelikož je velmi obtížné získat reprezentativní vzorek populace, je výskyt tohoto problému poměrně častý. Významnost tohoto problému je dále akcentována velmi častým využíváním různých typů průzkumů, a to jak v soukromých, tak také ve veřejnoprávních médiích.

Druhým případem je zaměňování jednotlivých typů středních hodnot mezi sebou nebo používání takových středních hodnot, které jsou pro vyjádření dané veličiny nevhodné či zavádějící. Vždy je nutné si uvědomit, který typ střední hodnoty je pro daný výsledek použit. Konstrukce jednotlivých středních hodnot je diametrálně odlišná a někdy samotní autoři daných textů mají problém s jejich správnou definicí a použitím.

Nejčastějším prvkem statistiky a prezentace dat, který je používán pro ovlivňování veřejného mínění, jsou však grafy. Zneužitelnost tohoto prvku statistiky plyne zejména z popularity, jednoduchosti a jeho zdánlivé srozumitelnosti. Proto je velmi

snadné pomocí několika drobných úprav, jako je například prostorový efekt v případě koláčového grafu či zkrácení intervalu osy Y ve sloupcovém grafu, vyvolat v příjemci informací zcela jiné závěry.

Posledním prvkem statistiky, který je k tvorbě a ovlivňování veřejného mínění využíván, je falešná korelace. Jedná se o dávání do vzájemných souvislostí zcela nesouvisející skutečnosti jenom na základě toho, že se dané veličiny vyvíjejí stejně nebo velmi podobně. V diplomové práci je uveden příklad, kdy mezi sebou silně koreluje počet smrtí oběšením s výdaji americké vlády na vědu. Vzájemná kauzalita obou veličin však již na první pohled neexistuje.

Hlavním prvkem diplomové práce je analýza dat sesbíraných prostřednictvím dotazníku. V období mezi 23. 4. 2017 a 14. 5. 2017 odpovědělo na otázky dotazníkového šetření celkem 219 respondentů. Samotný dotazník je rozdělen do třech sekcí. První sekce je tvořena otázkami zaměřenými na základní statistické pojmy. Druhá část dotazníku je zaměřena na schopnost respondentů interpretovat informace, které jsou podávány ne ve zcela řádné formě. Třetí část je zaměřena na osobní charakteristiky respondentů, mezi které je možné zařadit například pohlaví, status, věk či místo bydliště.

Z provedené analýzy osobních charakteristik dotázaných vyplynulo, že není možné vztáhnout závěry učiněné na základě sesbíraných dat k celé populaci. Důvodem je vychýlení tohoto vzorku, které se projevuje zejména u proměnné status a věková kategorie. Zastoupení jednotlivých skupin respondentů totiž nereflektuje skutečné zastoupení těchto skupin ve společnosti. Daná skutečnost však nevylučuje zodpovězení výzkumných otázek stanovených na začátku diplomové práce.

První výzkumná otázka se týkala identifikací faktorů, které ovlivňují znalost statistických pojmů. K tomuto účelu byl využit Chí-kvadrát test nezávislosti, jehož prostřednictvím došlo k prověřování závislosti znalosti statistických pojmů na jednotlivých osobních charakteristikách dotázaných. Kvůli menšímu počtu respondentů bylo nutné redukovat počet kategorií u některých proměnných do té úrovně, aby četnosti jednotlivých hodnot plnily předpoklady testování. Na základě analýzy bylo zjištěno, že znalost statistických pojmů závisí na věku, vzdělání a statusu respondentů. Naopak závislost na pohlaví a místě bydliště se nepodařila prokázat.

Proměnná věk byla v průběhu testování zredukována do třech kategorií. Z výsledků testů vyplynulo, že znalost statistických pojmů se snižuje se zvyšujícím se věkem respondentů. Významnou roli na tomto výsledku může hrát také skutečnost, že se v nejmladší skupině respondentů objevuje vysoká část studentů, kteří přicházejí s pojmy z této oblasti do styku častěji než ostatní věkové skupiny respondentů.

Hodnoty proměnné vzdělání byly za účelem splnění předpokladů testu sjednoceny do třech kategorií. V případě této proměnné platí přímá závislost, kdy zvyšující se stupeň vzdělání má pozitivní dopad na znalosti respondentů v oblasti statistiky. Nejúspěšněji si vedli vysokoškolsky vzdělaní respondenti, kdy celkem 83,33 % z nich dokázalo zodpovědět správně alespoň 3 z celkových 4 otázek zaměřených na problematiku statistických pojmů.

Poslední proměnnou, u které se podařil prokázat vliv na znalost statistických pojmů je status dotázaného. Z důvodu nízkého počtu dotázaných v některých skupinách došlo k jejich sjednocení do třech kategorií na studenty, výdělečně činné a ostatní. Nejlépe si v této otázce vedli studenti, kdy jich téměř 90 % z nich dokázalo správně odpovědět minimálně na 3 otázky z oblasti statistických pojmů. Naopak velmi špatných výsledků dosahuje skupina zahrnující důchodce, nezaměstnané a ostatní respondenty, kdy správně odpovídalo pouze 42,86 % dotázaných z této skupiny.

Třetí výzkumná otázka se týkala závislosti schopnosti interpretace dat obsažených v informačních sděleních na znalostech statistických pojmů. Na základě Chí-kvadrát testu nezávislosti, kdy výsledná p-hodnota byla menší než 0,001, byla prokázána závislost schopnosti správně porozumět informačnímu sdělení na znalosti statistických pojmů. Tento test je nosným testem celé práce a prokázání závislosti vyjadřuje důležitost znalostí z oblasti statistiky.

Analýzy provedené v rámci diplomové práce ukazují, jak snadné může zneužití statistiky k ovlivnění veřejného mínění být. Stačí posunout osu sloupcového grafu, natočit výseče koláčového grafu a informace na nás potom působí více svou formou než svým obsahem. Analýzy založené na dotazníkovém šetření dokazují, že i když schopnost odhalit nástrahy, které nám původci informací každý den chystají, závisí na spoustě faktorů, patří mezi ně také znalost statistických pojmů. Proto je důležité rozumět alespoň těm základním pojmům a ukazatelům statistiky a neustále mít na paměti, že existují pouze tři druhy lží – lež, zatracená lež a statistika.

Seznam literatury

A PROJECT OF THE GRADUATE INSTITUTE OF INTERNATIONAL STUDIES a GENEVA. Small arms survey 2007: guns and the city. Cambridge: Cambridge University Press, 2007. ISBN 9780521706544.

BARTOŠOVÁ, Jitka. Základy statistiky pro manažery. Vyd. 1. Praha: Oeconomica, 2006. 193 s. ISBN 80-245-1019-7.

Baumann, Martin. Jak lhat se statistikou? (výsledky průzkumu), 2017. Dostupné online na <https://jak-lhat-se-statistikou.vyplnto.cz>.

Countries with the most guns list has some surprises. CBC News [online]. 2016 [cit. 2017-06-11]. Dostupné z: <http://www.cbc.ca/news/world/small-arms-survey-countries-with-the-most-guns-1.3392204>

COONTZ, Stephanie. When Numbers Mislead. The New York Times [online]. 2013 [cit. 2017-04-16]. Dostupné z: <http://www.nytimes.com/2013/05/26/opinion/sunday/when-numbers-mislead.html>

ČERVENKA, Jan, KUNŠTÁT, Daniel (ed.). České veřejné mínění: výzkum a teoretické souvislosti. Praha: Sociologický ústav Akademie věd ČR, 2006. ISBN 80-7330-081-8.

Obyvatel České republiky přibylo. Český statistický úřad [online]. 2016 [cit. 2017-06-12]. Dostupné z: <https://www.czso.cz/csu/czso/obyvatel-ceske-republiky-pribylo>

DE VAUS, D. A. Analyzing social science data. Thousand Oaks, Calif.: SAGE, 2002. ISBN 0-7619-5938-6.

DISMAN, Miroslav. Jak se vyrábí sociologická znalost: příručka pro uživatele. 4., nezměn. vyd. Praha: Karolinum, 2011. ISBN 978-80-246-1966-8.

KOHOUT, Jaroslav. Veřejné mínění, image a metody public relations. Praha: Management Press, 1999. ISBN 80-7261-006-6.

FOTR, Jiří a Jiří HNILICA. Aplikovaná analýza rizika ve finančním managementu a investičním rozhodování. 2., aktualiz. a rozš. vyd. Praha: Grada, 2014. Expert (Grada). ISBN 978-80-247-5104-7.

FULLER, R. Buckminster. Critical path. New York, N.Y.: St. Martin's Press, c1981. ISBN 0312174918.

General Election: Trump vs. Clinton vs. Johnson vs. Stein. Real Clear Politics [online]. 2016 [cit. 2017-04-23]. Dostupné z: http://www.realclearpolitics.com/epolls/2016/president/us/general_election_trump_vs_clinton_vs_johnson_vs_stein-5952.html

GLASSER, Theodore Lewis. a Charles T. SALMON. Public opinion and the communication of consent. New York: Guilford Press, c1995. ISBN 0898624991.

Guns. Gallup [online]. 2016 [cit. 2017-06-11]. Dostupné z:
<http://www.gallup.com/poll/1645/guns.aspx>

HENDL, Jan. Kvalitativní výzkum: základní metody a aplikace. Praha: Portál, 2005. ISBN 80-7367-040-2.

HINDLS, Richard, Stanislava HRONOVÁ a Jan SEGER. Statistika pro ekonomy. 2. vyd. Praha: Professional Publishing, 2002. ISBN 80-86419-30-4.

HUK, Jaroslav. Výzkum veřejného mínění a mediální publikum. Vyd. 2., rozš. a přeprac. Praha: Univerzita Jana Amose Komenského, 2013. ISBN 978-80-7452-031-0.

CHALABI, Mona. Gun homicides and gun ownership listed by country [online]. 2012 [cit. 2017-06-11]. Dostupné z: <https://www.theguardian.com/news/data-blog/2012/jul/22/gun-homicides-ownership-world-list>

KAČEROVÁ, Eva a Libor MICHALEC. Příběh statistiky. Praha: Český statistický úřad, 2014. ISBN 978-80-250-2517-8.

Knowledge Doubling Every 12 Months, Soon to be Every 12 Hours. SCHILLING, David Russell. Industry tap [online]. 2013 [cit. 2016-12-04]. Dostupné z:
<http://www.industrytap.com/knowledge-doubling-every-12-months-soon-to-be-every-12-hours/3950>

KOZEL, Roman, Lenka MYNÁŘOVÁ a Hana SVOBODOVÁ. Moderní metody a techniky marketingového výzkumu. Praha: Grada, 2011. Expert (Grada). ISBN 978-80-247-3527-6.

LEIP, David. 1936 Presidential General Election Results. Dave Leip's Atlas of U. S. Presidential Elections [online]. 2016 [cit. 2017-04-23]. Dostupné z: <http://uselectionatlas.org/RESULTS/national.php?year=1936&f=0>

Lies, Damned Lies and Statistics. The University of York [online]. 2012 [cit. 2017-04-23]. Dostupné z: <https://www.york.ac.uk/depts/maths/histstat/lies.htm>

MIŠOVIČ, Ján. Od A do Z ve výzkumech veřejného mínění. Divišov: Orego, 2010. ISBN 978-80-86741-94-9.

NOVOTNÁ, Eliška, Jan NOVÝ a Martin MUSIL. Management public relations. Praha: Oeconomica, c2011. ISBN 978-80-245-1756-8.

Opinions on Gun Policy and the 2016 Campaign. Pew Research Center [online]. 2016 [cit. 2017-06-11]. Dostupné z: <http://www.people-press.org/files/2016/08/08-26-16-Gun-policy-release.pdf>

Průměrné mzdy - 3. čtvrtletí 2016. Český statistický úřad [online]. Praha, 2016 [cit. 2017-01-08]. Dostupné z: <https://www.czso.cz/csu/czso/cri/prumerne-mzdy-3-ctvrtleti-2016>

Počet starobních důchodců stoupá, průměrně pobírali 11 441 korun měsíčně [online]. Praha, 2016 [cit. 2017-05-20]. Dostupné z: <http://www.cssz.cz/cz/o-cssz/informace/media/tiskove-zpravy/tiskove-zpravy-2016/2016-11-3-pocet-starobnich-duxodcu-stoupa-prumerne-pobirali-11441-korun-mesicne.htm>

PŘEHLED VYBRANÝCH STATISTICKÝCH UKAZATELŮ Z AGEND ČSSZ. Česká správa sociálního zabezpečení [online]. Praha, 2015 [cit. 2017-05-20]. Dostupné z: http://www.cssz.cz/NR/rdonlyres/5B05D621-7022-460D-B140-5E077767E542/6398/vybrane_ukazatele_CSSZ_za_rok2014_20022015.pdf

ŘEZANKOVÁ, Hana. Analýza dat z dotazníkových šetření. 2. vyd. Praha: Professional Publishing, 2010. ISBN 978-80-7431-019-5.

SAUNDERS, M. N. K., Philip LEWIS a Adrian. THORNHILL. Research methods for business students. 5th ed. New York: Prentice Hall, c2009. ISBN 9780273716860.

SQUIRE, Peverill. Why the 1936 Literary Digest Poll Failed. Public Opinion Quarterly [online]. 52(1), 125- [cit. 2017-04-15]. DOI: 10.1086/269085. ISSN 0033362x. Dostupné z: <https://academic.oup.com/poq/article-lookup/doi/10.1086/269085>

SVOBODA, Václav. Public relations moderně a účinně. 2., aktualiz. a dopl. vyd. Praha: Grada, 2009. Expert (Grada). ISBN 978-80-247-2866-7.

SVOBODA, Helmut. Moderní statistika. Praha: Svoboda, 1977. 351 s. Členská knižnice.

ŠUBRT, Jiří. Kapitoly ze sociologie veřejného mínění: teorie a výzkum. Praha: Karolinum, 1998. ISBN 80-7184-522-1.

URBAN, Lukáš, Josef DUBSKÝ a Karol MURDZA. Masová komunikace a veřejné mínění. Praha: Grada, 2011. Žurnalistika a komunikace. ISBN 978-80-247-3563-4.

V České republice pracuje čtvrt milionu důchodců [online]. 2012 [cit. 2017-05-20]. Dostupné z: http://budoucnostprofesi.cz/aktuality.html/11_763-csu:-v-ceske-republice-pracuje-cvtrt-milionu-duxodcu

VIGEN, Tyler. Spurious Correlations. Hachette Books, 2015. ISBN 9780316339438.

Vývoj ekonomické aktivity obyvatelstva - 4. čtvrtletí 2011. Český statistický úřad [online]. 2012 [cit. 2017-06-12]. Dostupné z: <https://www.czso.cz/csu/czso/cri/vyvoj-ekonomicke-aktivity-obyvatelstva-4-ctvrtleti-2011-rwsr0r1lir>

WALTON, G. Charles. Policing public opinion in the French revolution: the culture of calumny and the problem of free speech. New York: Oxford University Press, c2009. ISBN 978-0-19-979580-2.

Zástupci tripartity jednali o zvýšení minimální mzdy. Česká strana sociálně demokratická [online]. Praha, 2015 [cit. 2017-04-23]. Dostupné z: <https://www.cssd.cz/media/tiskove-zpravy/zastupci-tripartity-zahajili-jednani-o-zvyseni-minimalni-mzdy/>

Přílohy

Příloha č. 1: Výsledky statistických testů

Tabulka 34 - Kontingenční tabulka proměnných znalost x pohlaví

Kontingenční tabulka – znalost – pohlaví				
		Pohlaví		Celkem
		Žena	Muž	
Znalost pojmů		6	6	12
Špatná	Četnost			
	Sloupcový podíl	4,72 %	6,52 %	5,48 %
Čás- tečná	Četnost	19	16	35
	Sloupcový podíl	14,96 %	17,39 %	15,98 %
Dobrá	Četnost	28	24	52
	Sloupcový podíl	22,05 %	26,09 %	23,74 %
Výborná	Četnost	74	46	120
	Sloupcový podíl	58,27 %	50,00 %	54,79 %
Celkem	Četnost	127	92	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 35 - Chí-kvadrát test nezávislosti (znalost x pohlaví)

Statistic	DF	Value	Prob
Chi-Square	3	1,5441	0,6722

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 36 – Kontingenční tabulka proměnných znalost x věková kategorie

Kontingenční tabulka – znalost x věk						
		Věková kategorie				Celkem
		Do 30 let	31-45 let	46-60 let	61 a více let	
Znalost pojmů		5	2	3	2	12
Špatná	Četnost					
	Sloupcový podíl	3,27 %	5,13 %	13,04 %	50,00 %	5,48 %
Částečná	Četnost	22	7	5	1	35
	Sloupcový podíl	14,38 %	17,95 %	21,74 %	25,00 %	15,98 %
Dobrá	Četnost	33	11	7	1	52
	Sloupcový podíl	21,57 %	28,21 %	30,43 %	25,00 %	23,74 %
Výborná	Četnost	93	19	8	0	120
	Sloupcový podíl	60,78 %	48,72 %	34,78 %	0,00 %	54,79 %
Celkem	Četnost	153	39	23	4	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 37 - Kontingenční tabulka proměnných znalost x věková kategorie po úpravě

Kontingenční tabulka – znalost x věk					
		Věková kategorie			Celkem
		Do 30 let	31-45 let	46 a více let	
Znalost		27	9	11	47
Ne	Četnost				
	Sloupcový podíl	17,65 %	23,08 %	40,74 %	21,46 %
Ano	Četnost	126	30	16	172
	Sloupcový podíl	82,35 %	76,92 %	59,26 %	78,54 %
Celkem	Četnost	153	39	27	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 38 – Chí-kvadrát test nezávislosti (znalost x věk)

Statistic	DF	Value	Prob
Chi-Square	2	7,3351	0,0255

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 39 - Kontingenční tabulka proměnných znalost x bydliště

Kontingenční tabulka – znalost – bydliště					
		Bydliště			Celkem
		Čechy	Morava	Slezsko	
Znalost pojmů		7	4	1	12
Špatná	Četnost				
	Sloupcový podíl	4,27 %	9,52 %	7,69 %	5,48 %
Částečná	Četnost	23	7	5	35
	Sloupcový podíl	14,02 %	16,67 %	38,46 %	15,98 %
Dobrá	Četnost	41	10	1	52
	Sloupcový podíl	25,00 %	23,81 %	7,69 %	23,74 %
Výborná	Četnost	93	21	6	120
	Sloupcový podíl	56,71 %	50,00 %	46,15 %	54,79 %
Celkem	Četnost	164	42	13	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 40 - Kontingenční tabulka proměnných znalost x bydliště po úpravě

Kontingenční tabulka – znalost – bydliště				
		Bydliště		Celkem
		Čechy	Morava a Slezsko	
Znalost pojmů		7	5	12
Špatná	Četnost			
	Sloupcový podíl	4,27 %	9,09 %	5,48 %
Částečná	Četnost	23	12	35
	Sloupcový podíl	14,02 %	21,82 %	15,98 %
Dobrá	Četnost	41	11	52
	Sloupcový podíl	25,00 %	20,00 %	23,74 %
Výborná	Četnost	93	27	120
	Sloupcový podíl	56,71 %	49,09 %	54,79 %
Celkem	Četnost	164	55	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 41 - Chí-kvadrát test nezávislosti (znalost x bydliště)

Statistic	DF	Value	Prob
Chi-Square	3	4,1833	0,2423

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 42 - Kontingenční tabulka proměnných znalost x vzdělání

Kontingenční tabulka - znalost - vzdělání							
		Vzdělání					Celkem
		Základní	SŠ	SŠ s maturitou	Vyšší odborné	Vysokoškolské	
Znalost		5	5	21	2	12	47
Ne	Četnost						
	Sloupcový podíl	50,00 %	71,43 %	19,27 %	22,22 %	14,29 %	21,46 %
Ano	Četnost	5	2	88	7	70	172
	Sloupcový podíl	50,00 %	28,57 %	80,73 %	77,78 %	83,33 %	78,54 %
Celkem	Četnost	10	7	109	9	84	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 43 - Kontingenční tabulka proměnných znalost x vzdělání po úpravě

Kontingenční tabulka - znalost - vzdělání					
		Vzdělání			Celkem
		ZŠ + SŠ bez maturity	SŠ + vyšší odborné	vysokoškolské	
Znalost		10	23	14	47
Ne	Četnost				
	Sloupcový podíl	58,82 %	19,49 %	16,67 %	21,46 %
Ano	Četnost	7	95	70	172
	Sloupcový podíl	41,18 %	80,51 %	83,33 %	78,54 %
Celkem	Četnost	17	118	84	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 44 - Chí-kvadrát test nezávislosti (znalost x vzdělání)

Statistic	DF	Value	Prob
Chi-Square	2	15,4964	0,0004

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 45 - Kontingenční tabulka proměnných znalost x status po částečné úpravě

Kontingenční tabulka – znalost – status						
		Status				Celkem
		Student	OSVČ	Zaměstnaný	Ostatní	
Znalost		10	2	27	8	47
Ne	Četnost					
	Sloupcový podíl	10,42 %	18,18 %	27,55 %	57,14 %	21,46 %
Ano	Četnost	86	9	71	6	172
	Sloupcový podíl	89,58 %	81,82 %	72,45 %	42,86 %	78,54 %
Celkem	Četnost	96	11	98	14	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 46 - Kontingenční tabulka proměnných znalost x status po konečné úpravě

Kontingenční tabulka – znalost – status					
		Status			Celkem
		Student	Výdělečně činný	Ostatní	
Znalost		10	29	8	47
Ne	Četnost				
	Sloupcový podíl	10,42 %	26,61 %	57,14 %	21,46 %
Ano	Četnost	86	80	6	172
	Sloupcový podíl	89,58 %	73,39 %	42,86 %	78,54 %
Celkem	Četnost	96	109	14	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 47 - Chí-kvadrát test nezávislosti (znalost x status)

Statistic	DF	Value	Prob
Chi-Square	2	15,4964	0,0004

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 48 - Kontingenční tabulka proměnných schopnost interpretace x znalost pojmů

Kontingenční tabulka – schopnost interpretace – znalost						
		Znalost				Celkem
		Špatná	Částečná	Dobrá	Výborná	
Schopnost interpretace		8	9	6	10	33
Špatná	Četnost					
	Sloupcový podíl	66,67 %	25,71 %	11,54 %	8,33 %	15,07 %
Částečná	Četnost	4	11	16	28	59
	Sloupcový podíl	33,33 %	31,43 %	30,77 %	23,33 %	26,94 %
Dobrá	Četnost	0	12	16	42	70
	Sloupcový podíl	0,00 %	34,29 %	30,77 %	35,00 %	31,96 %
Výborná	Četnost	0	3	14	40	57
	Sloupcový podíl	0,00 %	8,57 %	26,92 %	33,33 %	26,03 %
Celkem	Četnost	12	35	52	120	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 49 - Kontingenční tabulka proměnných schopnost interpretace x znalost pojmů po úpravě

Kontingenční tabulka – schopnost interpretace – znalost				
		Znalost		Celkem
		Ne	Ano	
Schopnost		32	60	92
Ne	Četnost			
	Sloupcový podíl	68,09 %	34,88 %	42,01 %
Ano	Četnost	15	112	127
	Sloupcový podíl	31,91 %	65,12 %	57,99 %
Celkem	Četnost	47	172	219

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Tabulka 50 - Chí-kvadrát test nezávislosti (schopnost interpretace x znalost pojmů)

Statistic	DF	Value	Prob
Chi-Square	1	16,7029	<0,0001

Zdroj: Dotazníkové šetření (2017), vlastní zpracování

Příloha č. 2: Dotazníkové šetření

Dobrý den,

rád bych Vás požádal o vyplnění dotazníku, jehož výsledky budou použity v mé diplomové práci. Cílem dotazníku je zjistit znalost statistických pojmů a schopnost interpretace zpráv, ve kterých se statistické údaje objevují.

Děkuji za Váš čas a těším se na zajímavé výsledky.

S pozdravem
Martin Baumann

1. Pepa dostal každý den v týdnu ke svačině různý počet jablek (viz tabulka).

Den v týdnu	Počet jablek
Pondělí	3
Úterý	1
Středa	3
Čtvrtek	3
Pátek	2

Aritmetický průměr z hodnot uvedených v této tabulce je roven číslu:

- a) Nevím
- b) 3
- c) 2,4
- d) 12
- e) 1
- f) 2

2. Pepa dostal každý den v týdnu ke svačině různý počet jablek (viz tabulka).

Den v týdnu	Počet jablek
Pondělí	3
Úterý	1
Středa	3
Čtvrtek	3
Pátek	2

Modus z hodnot uvedených v této tabulce je roven číslu:

- a) 3
- b) 2,4
- c) 12
- d) 1
- e) 2
- f) Nevím

3. Pepa dostal každý den v týdnu ke svačině různý počet jablek (viz tabulka).

Den v týdnu	Počet jablek
Pondělí	3
Úterý	1
Středa	3
Čtvrtek	3
Pátek	2

Medián z hodnot uvedených v této tabulce je roven číslu:

- a) 3
- b) 2,4
- c) 12
- d) 1
- e) 2
- f) Nevím

4. Rozptyl je:

- a) Hodnota, která se v daném souboru vyskytuje nejčastěji.
- b) Hodnota, která vyjadřuje, jak moc hodnoty souboru kolísají kolem střední hodnoty.
- c) Nevím.
- d) Součet všech hodnot souboru vydělený jejich celkovým počtem.

5. Titulek: Ceny nemovitostí jsou v roce 2017 na stejné úrovni jako v roce 2015!

Díky nízkým úrokovým sazbám se v roce 2016 cena nemovitostí zvýšila o 10 %, ale po vypuknutí hypoteční krize v roce 2017 naopak o 10 % klesla.

Správná odpověď je:

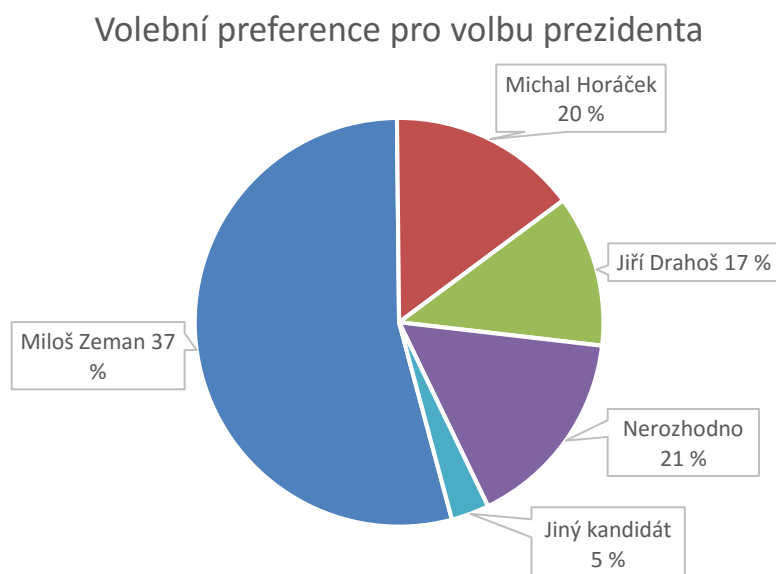
- a) Novinový titulek odráží skutečnost. Cena nemovitostí se opravdu nezměnila.
- b) Novinový titulek neodráží skutečnost. Ceny nemovitostí v roce 2017 jsou nižší než v roce 2015.
- c) Novinový titulek neodráží skutečnost. Ceny nemovitostí v roce 2017 jsou vyšší než v roce 2015.
- d) Nevím.

6. Agentura zabývající se průzkumem veřejného mínění na základě nedávného průzkumu zjistila, že se svým životem je spokojeno 85 % obyvatel České republiky. Průzkum probíhal v prodejně mobilních telefonů značky Apple, hotelu Imperial a v obchodě Hugo Boss a zúčastnilo se ho přes 1500 respondentů.

Správná odpověď je:

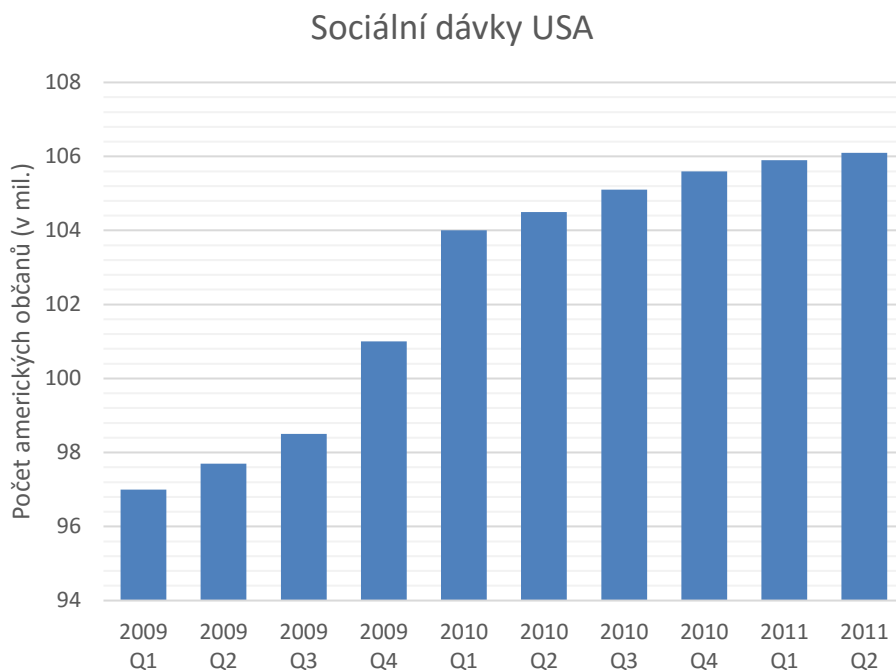
- a) Na základě provedeného průzkumu není možné stanovit skutečnou spokojenost obyvatel České republiky se svým životem.
- b) Provedený průzkum lze považovat za věrohodný. Se svým životem může být spokojeno 85 % obyvatel České republiky.
- c) Průzkum je smyšlený, žádný ze jmenovaných zařízení by nedovolilo pořádání průzkumu.
- d) Nevím.

7. Z níže uvedeného koláčového grafu vyplývá (více možností):



- a) Miloš Zeman by získal nadpoloviční většinu všech hlasů a obhájil by svůj mandát v prvním kole.
- b) Miloš Zeman by první kolo vyhrál, ale o budoucím prezidentovi by se rozhodlo až v kole druhém.
- c) Zatím nerozhodnuto je 21 % voličů.
- d) Neumím číst z grafů.

8. Z níže uvedeného sloupcového grafu vyplývá (více možností):



- a) Počet amerických občanů pobírajících dávky se zvýšil téměř trojnásobně.
- b) Počet amerických občanů pobírajících dávky se nezměnil, graf zobrazuje celkové výdaje v amerických dolarech.
- c) Počet amerických občanů pobírajících dávky narostl přibližně o 9 milionů.
- d) Neumím číst z grafů.

9. Jste:

- a) Žena
- b) Muž

10. Věková kategorie:

- a) Do 30 let
- b) 31–45 let
- c) 46–60 let
- d) 61 a více let

11. Jaké je Vaše nejvyšší dosažené vzdělání?

- a) základní

- b) středoškolské bez maturity
- c) středoškolské s maturitou
- d) vyšší odborné
- e) vysokoškolské

12. Jaký je Váš status?

- a) Student
- b) OSVČ
- c) Zaměstnaný
- d) Nezaměstnaný
- e) Důchodce
- f) Jiný

13. V jakém oboru jste zaměstnaný?

- a) Administrativa
- b) Obchod
- c) IT
- d) Veřejná správa
- e) Služby
- f) Průmysl
- g) Vlastní odpověď:

14. V jakém kraji bydlíte?

- a) Hlavní město Praha
- b) Středočeský
- c) Jihočeský
- d) Plzeňský
- e) Karlovarský
- f) Ústecký
- g) Liberecký
- h) Královehradecký
- i) Pardubický
- j) Vysočina
- k) Jihomoravský
- l) Olomoucký
- m) Zlínský
- n) Moravskoslezský