UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMTAICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

BAKALÁŘSKÁ PRÁCE
Lži a statistika

Vedoucí bakalářské práce:
Mgr. Jana Vrbková
Rok odevzdání: 2010

Vypracovala:
Martina Janáčková
ME, III. ročník

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně pod vedením Mgr. Jany Vrbkové s pomocí uvedené literatury a ostatních informačních zdrojů.

V Olomouci 12.4.2010 ----------

Martina Janáčková

Poděkování

Touto cestou chci především poděkovat svému vedoucímu práce Mgr. Janě Vrbkové za trpělivost, cenné rady, připomínky a čas, který mi věnovala při konzultacích.

Obsah

ÚVOD

„*Známe tři druhy lží: velkou lež, malou lež a statistiku*."
George Bernard Shaw

„*Jsou tři druhy lží – obyčejná, statistika, pravda*."
Hugo Werner

„*Jsou tři stupně lží: obyčejná lež, ďábelská lež a statistika*."
Disraeli

Tyto citáty nám jsou moc dobře známy a běžně se s nimi setkáváme při diskuzích s přáteli, zejména nad články novin, které se hemží číselnými údaji. Bohužel, ve spoustě případů je na výše uvedených výrocích mnoho pravdivého. Statistika může být brána i jako lež nebo se lží stát. Cílem této práce je upozornit na nejzávažnější chyby, které se vyskytují při vytváření statistik, ač jsou někdy vytvořeny neúmyslně. Sami dobře víte, jak lehce se dá manipulovat s míněním lidí pomocí statistik. Při čtení statistiky by se lidé měli ptát na několik základních otázek. Nesmí přijímat za pravdivé vše, co vidí v mediích. Na jaké otázky se ptát? To se dozvíte, v této práci. Inspirací k vytvoření tohoto textu mi byla kniha Damned lies and statistics [1] od autora Joela Besta.

Na začátku své knihy Best popisuje situaci, kdy byl požádán, aby se stal členem komise pro studentské diplomové práce. Při čtení jedné z diplomových prací ho zaujala citace: „Od roku 1950 se každý rok počet zastřelených dětí zdvojnásobil." Samozřejmě jako první ho napadlo, že se autor diplomové práce musel splést. Proto si Best tuto citaci ověřil v uvedeném zdroji. Šel do knihovny a vyhledal si článek, ze kterého student čerpal informace. Překvapilo ho, že citace je ve stejném znění jako ve studentské práci. Tato statistika se může docela jednoduše stát jednou z nejhorších a nejnepřesnějších sociálních statistik na světě. Co dělá tuto statistiku tak špatnou? Na vysvětlení budeme předpokládat, že v roce 1950 bylo zastřeleno jedno dítě. Další údaje jsou shrnuty v tabulce.

Tabulka č.1 Teoretický výpočet zastřelených dětí

Rok	Počet zastřelených dětí
1950	1
1951	2
1952	4
1953	8
...	
1960	$2^{10} = 1024$
1965	$2^{15} = 32768$
1980	2^{30} = asi 1 miliarda
1995	2^{45} = asi 35 bilionů

Přitom podle samotné FBI v roce 1965 potvrdila počet zavražděných lidí, dospělých i dětí, v celé zemi celkem na 9960 obětí. V roce 1995, kdy byl diskutovaný článek vydán, by roční počet obětí překročil 35 bilionů, což je nemožné. Proto jak bylo řečeno, je to jedna z nejnepřesnějších sociálních statistik. Best pátral, kde autor článku přišel k této úvaze. Autor mu odpověděl, že tato statistika vyšla od Fondu na ochranu dětí. V prohlášení tohoto fondu z roku 1994 stojí: „Počet zastřelených amerických dětí se zdvojnásobil od roku 1950.“ Tady už vidíme rozdíl ve formulaci, kdy fond tvrdí, že v roce 1994 bylo dvakrát více mrtvých dětí než v roce 1950. Autor článku toto tvrzení přeformuloval a tím pádem tomuto tvrzení dal zcela rozdílný význam.

Fond ve svých dokumentech tvrdí, že počet zastřelených dětí se zdvojnásobil od roku 1950 do roku 1994, což není tak dramatický nárůst jak by se mohlo zdát. Samozřejmě se také americká populace v té době zvětšovala. Můžeme předpokládat, že spousta hodnot, včetně počtu zastřelených dětí, bude růst už jen proto, že také populace roste. Daná statistika Fondu vyvolává další otázky: „Kdo počítal počet zastřelených dětí a jak?“ „Koho považujeme za dítě?“ - některé statistiky považují za dítě i osoby do 25 let. „Koho uvažujeme pod pojmem zastřelený?“- statistiky zastřelených lidí často zahrnují sebevraždy, nehody i vraždy. Tato statistika na autora článku zapůsobila a chtěl ji jen zopakovat. Místo toho ji přeformuloval a vytvořil „statistickou mutaci“.

Lidé ke zmutovaným statistikám obecně přistupují stejně jako k ostatním statistikám a často akceptují i ty nejnepravděpodobnější tvrzení bez jakýchkoli pochyb a otázek.

I editor časopisu, který dovolil publikovat tento článek, se vůbec neobtěžoval zvažovat význam článku s takto nesmyslným údajem. A lidé pak špatné statistiky používají dál. Student tuto nesmyslnou statistiku zkopíroval a použil ji ve své diplomové práci.

Na druhou stranu špatné statistiky jsou pro určité skupiny osob důležité. Mohou být použity (spíše zneužity) k tomu, aby vyvolaly vztek veřejnosti nebo strach. V některých případech mohou pokřivit naše chápání světa.

Kapitola 1

1.1 STATISTIKY V SOUVISLOSTI SE SOCIÁLNÍMI PROBLÉMY

Statistiky jsou často využívány v souvislosti se sociálními problémy. Debaty o sociálních problémech běžně vytváří otázky, které vyžadují statistické odpovědi: „Je tento problém obecně rozšířený?“ „Kolik lidí a jaké lidi problém ovlivňuje?“ „Zhoršuje se daná situace?“ „Co to bude stát, abychom se s problémem vypořádali?“ Přesvědčivé odpovědi na tyto otázky vyžadují důkazy a těmi obvykle myslíme čísla, měření, statistiky. Můžeme vůbec něco dokázat statistikou? Záleží na tom co slovem “dokázat“ máme na mysli.

Jestliže chceme vědět, kolik dětí bylo zastřeleno každý rok, nemůžeme prostě vyřknout číslo jen tak. Pokud takovéto číslo pro nás vypadá dostatečně přesné, můžeme ho považovat za velmi silné svědectví – důkaz. Řešením problému špatných statistik není jejich ignorování, nýbrž stát se lepšími soudci nad čísly, se kterými se setkáváme. Některé statistiky jsou špatné, ale některé jsou i velmi dobré a cenné. Statistiky potřebujeme – dobré statistiky, abychom mohli rozumně mluvit o sociálních problémech.

Sociální statistiky popisují společnost. Lidé, kteří nás na sociální statistiky upozorňují, mají důvod takto činit. Nevyhnutelně něco chtějí. Opakují a publikují statistiky k dosažení svých konkrétních cílů. Špatné statistiky pocházejí od konzervativců i liberálů, od bohatých společností a mocných agentur, ale také od advokátů chudých a bezmocných.

1.2 PRVNÍ SOCIÁLNÍ STATISTIKY

První „statistiky” měly za úkol ovlivnit debaty o sociálních záležitostech. Tento termín získal moderní význam – numerickou evidenci – kolem roku 1830 v čase, kdy newyorští reformátoři odhadovali, že město mělo na 10 000 prostitutek. Předchůdkyně statistiky se nazývala „politická aritmetika”. Tyto studie, které se objevily v 17.století, se většinou pokoušely spočítat velikost populace, zejména v Anglii a Francii. Analytikové se

pokoušeli spočítat porody, úmrtí a sňatky, protože věřili, že vzrůstající populace byla důkazem zdravého státu. Ti, kdo prováděli tyto numerické metody, je začali nazývat *statistiky.*

Tím, že badatelé opakovaně shromažďovali a analyzovali data, začali v nich vidět určité vzorce. Rok od roku objevovali, že čísla porodů, úmrtí a dokonce sňatků zůstala relativně stabilní. Tato stabilita naznačovala, že to, co se stalo ve společnosti, záleželo na více než jen pouhém jednání vlády.

1.3 SOCIALNÍ STATISTIKA V 19. STOLETÍ

Na začátku 19. století se sociální pořádek zdál obzvláště ohrožen. Města byla větší, než kdy předtím, ekonomiky se začínaly industrializovat a revoluce v Americe a Francii prokázaly, že politická stabilita se nedá pokládat za samozřejmost. Různé agentury začaly shromažďovat a publikovat statistiky. Spojené státy a několik evropských států zavedly pravidelné sčítání lidu ke shromáždění populační statistiky. Soudy, věznice, policie začaly zaznamenávat počty přestupků a trestných činů, lékaři počty pacientů, učitelé počítali studenty a tak dále.

Reformátoři usilující o řešení sociálních problémů 19. století – chudoba, nemoci, upadající mravnost, práce dětí, pracovní síla továren a související oslabení zemědělských prací – pokládali statistiky za užitečnou metodu dokládající rozsah a krutost utrpení. Statistiky poskytly oběma – vládě i reformátorům – důkazy. Čísla udávala jistou přesnost - místo debaty o prostituci, jakožto nejasně definovatelném problému, reformátoři začali uvádět specifická číselná tvrzení.

V průběhu 19. století se statistiky staly autoritativním způsobem jak popsat sociální problémy. Začal růst respekt k vědě a statistiky nabízely možnost jak přenést autoritu vědy do debat nejen o sociální politice. Ve skutečnosti právě tohle byl jeden z cílů prvních statistiků – chtěli studovat společnost na základě sčítání a výsledná čísla uplatnit při ovlivňování sociální politiky. Uspěli. Široké přijetí statistik se rozšířilo jako prostředek

k měření nejen sociálních problémů. Od začátku 19. století až dosud, mají statistiky dva účely – jeden je veřejný, druhý je často skryt.

Jejich veřejný účel je podat přesný, pravdivý popis stavu společnosti. Ale lidé také používají statistiku pro podporu specifických názorů na sociální problémy. Čísla tak mají i politický účel, který je často skryt za tvrzením, že čísla - jednoduše protože jsou čísly - musí být správná. Je ale naivní považovat čísla za správná bez toho, abychom věděli, kdo je používá, proč a jak se k nim přišlo.

Jeden příklad z historie prostituce z USA z období 19. století. Reformátoři ji nazývali „sociální ďábel“ a varovali, že mnohé ženy se samy prodávaly. Kolik žen? Jen pro město New York tu bylo několik odhadů. V roce 1833, reformátoři publikovali posudek dokládající, že v New Yorku bylo „ne méně než 10 000 prostitutek“ (rovnající se 10% ženské populace města). V roce 1866 newyorský metodistický biskup prohlašoval, že ve městě bylo více prostitutek (11 – 12 000) než metodiků. Další odhady se šplhali až k 50 000. Tito reformátoři doufali, že jejich zprávy o šíření prostituce podnítí autority k okamžitému jednání, ale městští úředníci spíše napadali odhady těchto reformátorů. Různá policejní a soudní vyšetřování odhalila mnohem nižší čísla. Jedna policejní zpráva z roku 1872 počítala pouze s 1 223 prostitutkami (v té době byla populace žen v New Yorku zhruba půl milionu obyvatel). Historikové vidí jasný vzorec v těchto cyklech konkurenčních statistik. Ministři a reformátoři inklinovali k přehánění, kdežto policejní úřednicí prostituci naopak podceňovali. Tyto matoucí statistiky připomínají i jiné, spíše současné debaty.

Během prezidentského období Ronalda Reagana aktivisté prohlašovali, že 3 miliony Američanů jsou bez domova, kdežto Reaganova vládní administrativa trvala na tom, že aktuální čísla lidí bez domova se přibližovala 300 000, jedné desetině toho, co uváděli aktivisté. Jinými slovy aktivisté tvrdili, že bezdomovectví je velký problém, který vyžaduje dodatečné vládní programy. Na druhé straně vládní administrativa prohlašovala, že nové programy nejsou třeba, jelikož čísla nebyla tak vysoká, aby vyžadovala speciální řešení. Každá ze stran prezentovala svou statistiku, která ospravedlňovala její doporučený postup a každá také kritizovala čísla strany druhé. Aktivisté se vysmívali vládním počtům jako pokusu skrýt viditelný problém do ústraní, zatímco vláda si stále za tím, že čísla

aktivistů jsou nereálně zvětšená. Statistiky se v takovém případě stanou zbraní v politickém boji o sociální problémy a politiku.

1.4 MÉDIA A SOCIÁLNÍ PROBLÉMY

Sociální problémy mají své příčiny v uspořádání společnosti. Pokud se ženy obrátí k prostituci nebo někteří lidé nemají domov, domníváme se, že společnost selhala (ačkoli můžeme i nesouhlasit – selhání způsobuje nedostatek práce, nepředávání dostatku morálních zásad dětem, nebo cokoli jiného). Sociologové mluví o tom, že sociální problémy bývají vykonstruované – to je vytvořené či shromážděné pomocí akcí aktivistů, úředníků, pomocí médií a dalších lidí poutající pozornost ke konkrétním problémům. Sociální problém je nálepka, jíž dáváme nějakému sociálnímu stavu a je to právě tato nálepka, která promění stav, jenž pokládáme za samozřejmý, v něco, co považujeme za znepokojující.

Když začneme uvažovat o prostituci nebo bezdomovectví jako o problému, reagujeme tím na kampaně reformátorů usilujících o vyburcování našeho zájmu v této oblasti. Hlavní role často hrají aktivisté, kteří se snaží upozornit ostatní na problém. Pozornost lidí přitahují protestními demonstracemi, zapojují média, nabírají nové členy, apelují na úředníky, aby něco udělali, a tak dále. Často si také zajistí podporu expertů – lékařů, vědců, ekonomů, kteří by měli mít dostatečnou kvalifikaci hovořit o důležitosti nějakého problému. Experti můžou provést výzkum na dané téma a přednést své výsledky. Užívání expertů působí velmi autoritativně a také média často na jejich názory spoléhají, jakožto na kvalitní prostředek ke tvorbě článků.

Ne všechny společenské problémy jsou propagovány bojem nezávislých aktivistů – tvorba nových problémů je také někdy práce mocných institucí a organizací. Vládními úředníky počínaje, kteří se snaží přitáhnout pozornost k nadcházejícím volbám, až k anonymním byrokratům konče, kteří se snaží přesvědčit, že rozšíření působnosti jejich agentur povede k vyřešení (nejen) sociálního problému. Veřejné i soukromé organizace si mohou dovolit najmout experty k provedení výzkumu, sponzorovat a podporovat aktivisty a publikovat jejich poznámky tak, aby zaujaly pozornost médií.

Snaha vytvořit nebo propagovat sociální problémy, obzvláště pokud přitahují pozornost, mohou inspirovat opozici. Když se nám poprvé dostane do podvědomí záležitost sociálního problému (nejspíše z reportáže televizních novin), dostaneme pravděpodobně jeden či dva příklady (např. video o lidech žijících na ulici) a nakonec statistický odhad (počet lidí bez domova). Většinou dostaneme vysoké číslo, jelikož právě to nás varuje, že problém je nezanedbatelný a vyžaduje naši pozornost, zájem, akci.

Média mají statistiky velmi v oblibě, neboť čísla se zdají být "faktem". Také aktivisté dožadující se pozornosti zjišťují, že tisk čísla přímo potřebuje. Reportéři se opírají o odhady, jaká je závažnost problému – kolik lidí je postiženo, kolik to stojí, apod. Média upřednostňují rušivé statistiky o velkých problémech, protože velké problémy znamenají zajímavější, přesvědčivější zprávy. Stejně tak jako výzkum expertů se zdá důležitější, pokud se jedná o závažný problém. Tyto skutečnosti vedou lidi k tomu, aby publikovali statistiky pro podporu svých pozic, důvodů a zájmů. Existuje výraz, který zachycuje tuto tendenci: „Čísla možná nelžou, ale lháři číslují." Lidé debatující o sociálním problému si vybírají statistiky k podpoře svých názorů.

Advokáti zastávající regulaci zbraní budou pravděpodobněji podávat zprávu o zastřelených dětech, zatímco oponenti budou počítat obyvatele, kteří vlastní zbraň odůvodněně ke své ochraně. Obě čísla mohou být správná, ale většina lidí debatujících o regulaci zbraní, bude prezentovat tu statistiku, která podpoří jejich postoj. Nejvíce tvrzení stahujících pozornost k novému problému míří k přesvědčení nás všech – členů široké veřejnosti. Pokud se společnost přesvědčí, že prostituce nebo bezdomovectví je závažný problém, pak je pravděpodobnější, že se něco pro řešení této situace udělá. Veřejnost tíhne k souhlasu, jednáme se statistikami jako s pravdou. Částečně proto, že jsme číselně negramotní.

Negramotnost v počtech je matematický ekvivalent analfabetismu. Je to „neschopnost" jednat s pojmy jako je číslo. Pokud se společnost chystá živit bezdomovce, mít přesné číslo je stejně tak potřeba, jako při plánu večeře pro 3 nebo 30 hostů. Číselná negramotnost způsobuje, že mnoho statistických tvrzení o sociálních problémech nedostane tolik pozornosti, kolik si zaslouží. To není jednoduše proto, že číselně negramotná společnost je manipulovaná obhájci, kteří propagují nesprávné statistiky.

Statistiky o všech problémech často opravdově pocházejí od lidí, kteří smýšlejí dobře, ale sami jsou číselně negramotní - možná plně nepochopí důsledky toho, co říkají.

Podobně i média nejsou imunní k číselné negramotnosti. Reportéři obvykle opakují čísla získaná od svých zdrojů, aniž by o nich kriticky zapřemýšleli. Vědí, že vysoká čísla značí velký problém. Aktivisté produkují velký odhad a tisk ho jednoduše otiskne. Široká veřejnost (většina trpí alespoň mírnou číselnou negramotností) tíhne k akceptování čísel bez jakýchkoli otázek. Jedním z důvodů, proč nekriticky akceptujeme statistiky je, že předpokládáme, že čísla pochází od expertů, kteří vědí, co dělají. Data můžou pocházet od vlády – míra kriminality, míra nezaměstnanosti jsou oficiální statistiky. Tady se objevuje přirozená tendence brát tato čísla bez jakýchkoli otázek jako přímá fakta.

Příkladem může být patolog. Člověk, který rozhoduje, zda úmrtí je sebevraždou či ne. Může to být jasné, mrtvý zanechal dopis, ve kterém uvádí svůj úmysl spáchat sebevraždu. Ale i přesto se musí shromáždit důkazy, které na sebevraždu ukazují – např. že trpěl depresemi apod. Může dojít ke dvěma potencionálním omylům. Prvním omylem může být, že patolog dá úmrtí nálepku "sebevražda", i když ve skutečnosti tu byla jiná příčina (může to být nafingované). Opačným omylem může být, když patolog přiřadí jinou příčinu úmrtí tomu, co byla ve skutečnosti sebevražda. Pozůstalí členové rodiny se mohou dožadovat, aby patolog prohlásil sebevraždu za nehodu (z důvodu zahanbení, životní pojistka). Jinak řečeno, oficiální záznamy sebevražd odráží patologovo rozhodnutí o příčinách smrti v situacích, kdy jsou okolnosti nejasné. Stává se, že různí patologové vyhodnotí případy různě. Patologové jsou více zaujatí schopností ospravedlnit svá rozhodnutí v individuálních případech než celkovou statistikou plynoucí z jejich rozhodnutí. Příklad záznamů sebevražd ukazuje na to, že všechny oficiální statistiky jsou produkty – a často vedlejší produkty - rozhodnutí různých úředníků, nejen patologů, ale také prostých úředníků, kteří vyplňují formuláře a zakládají je.

Zjednodušeně řečeno, i oficiální statistiky jsou sociálními produkty formované lidmi a organizacemi, které je vytvářejí. Mluvíme o statistikách jako o faktech, která jednoduše existují. Všechny statistiky jsou vytvářeny lidskými činy. Lidé musí rozhodnout, co počítat a jak to počítat, lidé musí provést sčítání a kalkulování a lidé musí interpretovat statistické výsledky, tj. popsat, co čísla znamenají. Měli bychom být schopni rozlišit špatné statistiky od těch dobrých.

Kapitola 2

2.1 ZDROJE ŠPATNÝCH STATISTIK

Každý den se v tisku setkáváme s čísly, která něco vypovídají o naší společnosti. Příkladem může být výrok odpůrců tabákových výrobků, kteří přisuzují 400 000 úmrtí za rok vlivu kouření. Tisk obecně rád prezentuje tato čísla jako fakt – nepopiratelnou skutečnost. Někdo tato čísla musel vyprodukovat. Jak? Existují snad nějaké lékařské kapacity, jež určily, za která úmrtí způsobená plicní rakovinou mohou cigarety a za která jiné důvody, například znečištěné ovzduší?

I ti nejlepší tvůrci statistik nemohou zaručit naprostou správnost průběhu a bezchybnost výsledků statistického rozboru (analýzy). V první řadě se jedná o výskyt chybných, duplicitních (nadbytečných) nebo chybějících dat. Při velkém počtu údajů je vysoká pravděpodobnost výskytu těchto chyb. I extrémním hodnotám se musí věnovat zvláštní pozornost, neboť vzniká podezření na důsledek hrubé chyby při zjišťování. Často se požaduje, aby daná (extrémní) hodnota byla při další analýze vypuštěna. V druhé řadě se jedná o zaokrouhlovací chyby nebo chyby vzniklé odřezáním „nedůležitých“ čísel ze zpracovávaných údajů. Další kapitolou jsou subjektivní omyly statistiků, spočívající v použití nevhodného metodického aparátů. [8]

Musíme pamatovat na to, že lidé, kteří propagují sociální problém, chtějí přesvědčit ostatní o „své pravdě“ a k tomu využívají i statistiku. Způsob, jakým tito lidé vytvářejí statistiky, je většinou špatný – jimi poskytnutá čísla mohou být často snad jen trochu víc než jen subjektivní odhady a spekulace. Jejich výpočty mohou být výsledky nedostatečných definic (vymezení pojmu), špatných měření nebo nevhodných příkladů.

2.2 TEMNÁ FIGURA

Ve většině případů, kdy aktivisté propagují své statistiky, používají „hádání“. Pokud je znepokojující sociální problém společností či vládou ignorován, obvykle tu nejsou žádné přesné záznamy, které by byly základem pro dobré statistiky. Nejen kriminalisté proto používají výraz „temná figura“, odkazující na poměr kriminálních činů, které se neobjeví ve statistikách kriminality. Občané hlásí všechny kriminální činy na policii, policie vše zaznamená a později se tyto údaje stávají podklady pro výpočet míry kriminality. To je jen teorie. Samozřejmě ne všechny trestné činy jsou nahlášeny (kvůli strachu ze msty, z prozrazení identity nebo jen proto, že si lidé myslí, že by jim policie stejně nepomohla). Může se také stát, že policie nezaznamená všechny trestné činy. Rozdíl mezi číslem oficiálně zaznamenaným a skutečným počtem kriminálních činů je tzv. *temná figura*.

Každý sociální problém má temnou figuru. Jak velká je temná figura? Když se poprvé dozvídáme o novém problému, ke kterému jsme nikdy předtím nevzhlíželi, můžeme o temné figuře uvažovat jako o problému. Nevíme, kolik nenahlášených případů existuje. Na druhou stranu se v jiných případech může jednat jen o „malou“ temnou figuru, např. u vražd. Z důvodu toho, že mrtvá těla se snadno dostanou do pozornosti policie.

Odhady aktivistů jsou většinou přeceňováním daného problému. Reportéři většinou žádají od aktivistů nějaká čísla, neboť sami nejsou schopni nalézt jiné, přesnější statistiky. Díky tomu, že se dané číslo objevuje ve zprávách, v proslovech politiků, odborných časopisech, lidé ztratí pojem o původu tohoto čísla. S tím samozřejmě souvisí zkrášlování statistik, kdy lidé mění význam statistik.

2.3. DEFINOVÁNÍ

Každý problém zahrnuje i jeho definici. Žádná z definic není dokonalá. Jsou tu dva hlavní způsoby, jak takovou definici můžeme pokazit. První způsob je, že definice je příliš široká, zahrnuje klamná pozitiva (tzv. falešné pozitivní případy, které by měly být vynechány). Druhý způsob je přesný opak, definice je příliš omezená, vylučuje klamná negativa (tzv. falešně negativní případy, jež by měly být zahrnuty). Již samotná definice problému může ovlivnit statistiky – číselné vyjádření daného problému. Čím bude definice obecnější, tím jednodušeji budeme moci ospravedlnit vysoké odhady.

Pokud někdo oznámí, že 16% Čechů [3] je negramotných, je velmi důležité ptát se, jak je v tomto případů definovaná negramotnost. Můžeme předpokládat, že negramotnost znamená, že člověk neumí vůbec číst a psát. Ale můžeme mít na mysli i částečnou negramotnost (neschopnost číst noviny, mapy, vyplnit dotazník pro získání pracovního místa či daňový formulář). Kdyby byla negramotnost přesně definovaná, bude identifikována možná u mnohem méně lidí a proto bude přesnější vymezení pojmu produkovat mnohem nižší statistické odhady než obecná definice. Z toho vyplývá, že čím je pojem definován obecněji, tím je statistika vyšší.

2.4 MĚŘENÍ

Každá statistika založená na něčem více než jen pouhém hádání, vyžaduje určité počítání, zjišťování. Definice zkoumaného problému specifikují, co bude počítáno. Jedním z nejčastějších způsobů jak např. příslušné orgány měří sociální nespokojenost je *průzkum*. Průzkumy zahrnují dotazování lidí, sčítání jejich odpovědí a načrtnutí všeobecného závěru založeného na výsledcích. V další části této práce se budu vytváření průzkumů podrobněji věnovat. Obhájci nějakého problému, kteří si mohou dovolit sponzorovat své vlastní průzkumy, mohou vymodelovat výsledky průzkumu „k obrazu svému“. Formulují otázky

tak, aby účastníci průzkumu odpovídali požadovaným způsobem. Také pořadí otázek může ovlivnit odpovědi respondentů. Nejen, že tvůrci takovýchto průzkumů ovlivňují otázky a jejich pořadí, ale také jen oni sami se můžou rozhodnout, jak budou interpretovat výsledky, případně, které výsledky veřejnosti poskytnout a které ne.

2.5 VÝBĚR Z POPULACE

Prakticky všechny sociální statistiky zahrnují zevšeobecňování na základě výběru z populace. Pro odborníka - statistika jsou zkoumané případy výběrem, který dostatečně dobře reprezentuje větší populaci všech případů. Nejen malé, ale i národní průzkumy obvykle vyzpovídají 1000 až 2000 lidí. Je zřejmé, že čím bude vzorek menší, tím méně bude odrážet situaci v celé populaci. Ani příliš velký vzorek nemusí být dobrá volba. Mnohem důležitější je, aby vzorek byl reprezentativní. Dobrý vzorek přesně odráží populaci. V ideálním případě lidé odhadující statistiku znají přesný rozsah populace, kterou chtějí studovat a umí z této populace sestavit reprezentativní vzorek. Velmi často je nemožné definovat celou populaci a určení reprezentativního výběru z populace je velmi nákladné a zdlouhavé.

Kapitola 3

DOBRÉ STATISTIKY A ŠPATNÉ STATISTIKY

Když narazíme na novou statistiku, musíme se ptát na několik otázek.

- *Kdo vytvořil tuto statistiku?* Každá statistika má své autory, své tvůrce. Někdy číslo pochází od jednotlivce. Jindy jsou to velké organizace či agentury. Musíme se ptát na otázky typu: „Nepochází čísla od aktivistů, kteří se jen snaží získat pozornost a vyburcovat zájem o daný problém?“ „Není toto číslo publikováno médii jen ve snaze dokázat, že tento problém je dostatečně zajímavý pro čtenáře (= pro noviny)?“
- *Proč byla tato statistika vytvořena*? Musíme si uvědomit, že lidé vytvářející statistiky jsou často zaujati tím, co čísla ukazují, používají čísla jako nástroj k přesvědčování.
- *Jak byla tato statistika vytvořena?* Musíme se zejména ptát, jak autoři ke statistice přišli. Dobré statistiky jsou více než na hádání založeny na odborně vedeném a poctivém průzkumu.

3.1 ZNAKY DOBRÉ STATISTIKY

I. Dobré statistiky jsou založeny na jasné a srozumitelné definici zkoumaného problému. Každá statistika musí dostatečně přesně definovat svůj předmět.

II. Dobré statistiky vznikají díky jasnému a dobře vedenému měření (hodnocení). Žádná statistika není dokonalá, ale některé vady jsou závažnější než jiné. Lidé, kteří produkují číselné údaje (statistiky), by měli být ochotni vysvětlit a doložit, jak problém měřili.

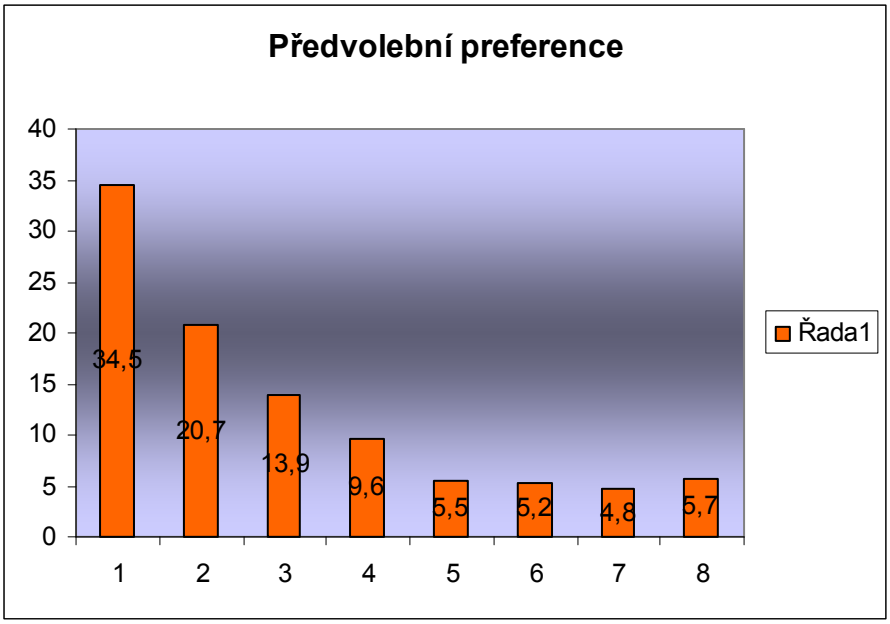
III. U statistiky by měla být přesně specifikována samotná populace a způsob získání výběru z této populace, u kterého se daný problém zkoumal.

Zde je uveden příklad kvalitní i nekvalitní statistiky.

„Správná“ statistika

Terénní sběr dat proběhl ve dnech 1.2. – 2.3. 2010 stratifikovaným adresným náhodným výběrem. Respondenti tvoří reprezentativní vzorek populace ČR dle údajů ČSÚ. Předvolební průzkum se účastnilo 1051 respondentů starších 18 let. Metodologie sběru dat – osobní rozhovor tazatele s respondentem, který zaznamenává odpovědi do elektronického dotazníku (notebooku). Do výsledků jsou zahrnuti oprávnění voliči, kteří by šli k volbám a jsou rozhodnuti, kterou stranu by volili.

Obrázek č.1 Předvolební průzkum podle agentury Median

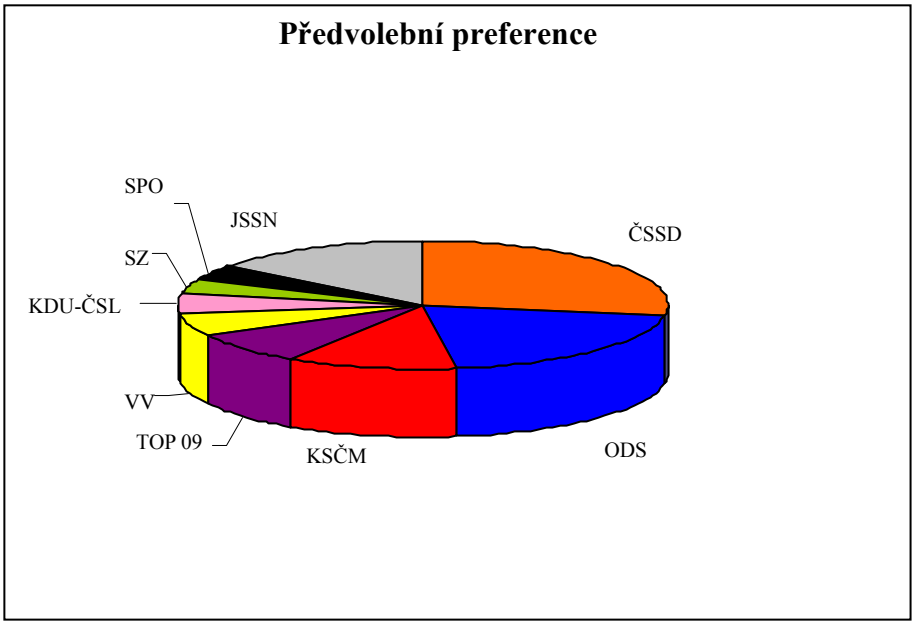


Zdroj: [13]

Vidíme, že je přesně stanovená definice průzkumu, způsob sběru dat. I populace je konkrétně vymezena. Je uveden i přesný počet respondentů.

„Špatná“ statistika

Obrázek č.2 Předvolební průzkum dle [14]



27.9% ČSSD Česká strana sociálně demokratická

20% ODS Občanská demokratická strana

11.3% KSČM Komunistická stran Čech a Moravy

7.7% TOP 09 TOP 09

6.2% VV Věci veřejné

4.8% KDU-ČSL Křesťanská a demokratická unie - Československá strana lidová

3.8% SZ Strana zelených

3.6% SPO Strana Práv Občanů

14.7% JSNN Jiné strany, nerozhodnutí a nevoliči.

Nikde není popsána metodologie sběru dat, neznáme ani přesný počet respondentů. Neznámé přesnou otázku na respondenty.

Kapitola 4

4.1 MUTANTNÍ STATISTIKY

Čísla, dokonce i správná čísla, mohou být nepochopena nebo špatně interpretována. Jejich význam může být otočen, překroucen nebo jinak pozměněn. A právě tyto změny vytvářejí statistiky „zrůdy“ (mutanty) – zdeformované verze originálních čísel. Některé mutantní statistiky jsou úmyslné pokusy pozměnit informace za účelem učinit určitá tvrzení přesvědčivější. Tohle se většinou stává, pokud se mutace objeví v rukou velkých institucí, které překroutí informace vyhovující jejich zájmu. Jakmile někdo představí mutantní statistiku, je tu velká šance, že čtenáři nebo posluchači ji přijmou a budou ji dále rozšiřovat. Zejména, je-li statistika vyřčena politickým lídrem, respektovaným komentátorem či jinou mediálně známou osobou. Daná čísla se stávají hůře zpochybnitelná, neboť lidé o nich slyšeli v mediích a předpokládají, že jsou správná.

Jak a proč se mutace objevuje? Začíná to *nevhodným zevšeobecňováním* statistiky. Poté se statistika mění *transformací* – vezme se číslo, které něco znamená a interpretuje se úplně jinak. A další fáze mutace je *zmatení* – transformace, které obsahují nepochopení významu komplikovanější statistik.

Zevšeobecňování

Ne vždy je v lidských silách spočítat například všechny případy určitého problému. Místo toho se shromáždí určitá evidence (ze vzorku - výběru) a získané závěry se poté zevšeobecní na všechny případy. Základní principy správného zevšeobecňování spočívají ve správné definici (problému, populace), jasném měření a reprezentativním vzorku populace. Mutantní statistika vzniká porušením jednoho z uvedených principů.

Transformace – změna významu statistiky

Transformace statistiky se většinou týká těch statistik, které jsou často opakovány. Obsahuje změnu významu - statistika X se změní na statistiku Y. Transformace může být i

neúmyslná, jestliže je způsobena nedbalým a nepřesným vyjádřením statistiky. V takovém případě se lidé snaží zopakovat statistiku, ale neúmyslně přeformulují tvrzení a tím pádem získá dané číslo úplně jiný význam. Na druhou stranu někteří lidé úmyslně „vylepšují“ zveřejněné statistiky. Takové statistiky se opakují velmi dobře, neboť většinou obsahují dramatická, znepokojující čísla. Je snadné udělat transformační chyby, ale obtížnější je jejich napravování. Stačí, aby jediný člověk nepochopil statistiku a šířil ji dál v pozměněné formě.

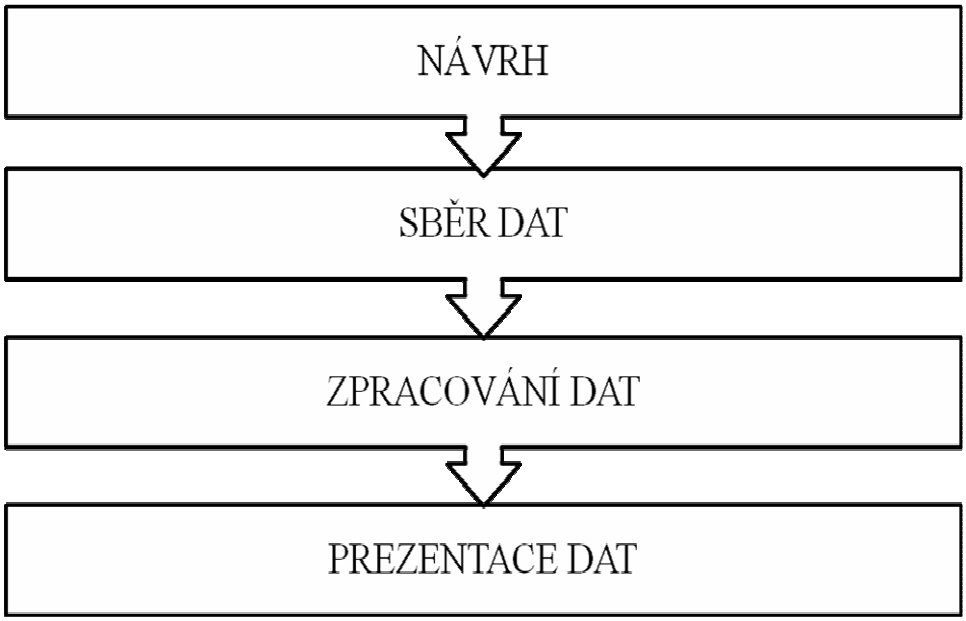
Zmatek – překroucení komplikovaných statistik

Zde nastává problém se zkomolením statistik z důvodu jejich složitosti. Jednoduchým statistikám většina lidí správně porozumí. Ale spousty sociálních, ekonomických a dalších problémů obsahují složité řetězce statistik, které mohou být pochopené pouze díky složitým modelům, které vyvinuli experti a také jim často jen oni sami dokáží porozumět.

Kapitola 5

5. 1 OBECNÉ SCHÉMA PRŮZKUMU

Obrázek č. 3 Obecné schéma dílčích stadií průzkumného projektu



5.2 ZÁKLADNÍ SOUBOR. ÚPLNÉ, NEÚPLNÉ ŠETŘENÍ.

Prvním krokem jakéhokoliv průzkumného projektu je formulování zkoumaného problému a také stanovení cíle daného průzkumu. Každý průzkum vyžaduje určité náklady finanční i časové, které by měly odpovídat dosaženému efektu – získaným informacím.

Jak by měl vypadat základní soubor? Data, která jsou získána „v terénu“ určitým šetřením se nazývají *primární*. Nositelem primárních informací v sociálních průzkumech je konkrétní osoba a základní soubor obvykle tvoří skupina osob, populace. Jednotkou šetření mohou být i domácnosti, instituce či firmy. Soubor, na který hodláme výsledky zkoumání zobecnit, a soubor, z něhož pořizujeme vzorek, musí být identický [2]. Například

rozhodneme-li se některé jednotky při vybírání vypustit (jsou těžko dostupné, apod.), je nutno tuto skutečnost zohlednit při zobecňování.

Víme, že v mnoha případech nemůžeme studovat celou populaci, která nás zajímá. Proto se musí přesně vymezit populace, které se průzkum bude týkat. Též se může uvažovat populace čistě hypoteticky, kdy je vymezena pouze vlastnostmi osob, které se řadí do zkoumané populace (např. žena v reprodukčním věku, běloška, krevní skupina A, Rh faktor negativní). A právě tady se musíme rozhodnout, kterou metodu šetření si vybereme, úplnou či neúplnou (výběrovou).

Úplné šetření spočívá v tom, že se prošetří všechny jednotky souboru (cenzus). Příkladem takovéhoto šetření je sčítání lidu, domů a bytů k určitému okamžiku, ale i demografické jevy jako je smrt, či narození. Tato metoda je velmi časově i finančně náročná. V některých případech je i nemožná, neboť celek (základní soubor, populace) bývá velmi rozsáhlý.

Při *neúplném šetření* se sledovaná vlastnost zjišťuje jen u některých prvků (statistických jednotek), tvořící výběr. Informace jsou tedy zjišťovány pouze na vybrané části celku (populace) a samotný výběr může výsledky průzkumu značně ovlivnit. Často se stává, že je to jediná proveditelná metoda. Těchto výběrů z jednoho základního souboru může být mnoho a zjištěné výsledky budou mít proměnlivou vypovídací hodnotu, proměnlivost (variabilitu). Z pohledu celého souboru poskytuje výběrové šetření pouze přibližné charakteristiky celku – zkoumané populace, pro kterou chceme zjištěné závěry zobecnit. Pokud je výběr tvořen vhodně, např. náhodným výběrem, můžeme na jeho základě získat určitou představu o celku. Z výběrového souboru nelze přesně určit parametry populace (centrální, tj. střední hodnotu, variabilitu - směrodatnou odchylku, apod.) charakterizující celek, ale pouze jejich odhady (výběrový průměr, výběrovou směrodatnou odchylku).

Úplné šetření má tedy oproti neúplnému několik výhod:

a) Úplné šetření poskytuje informace o celém souboru i o jednotlivých prvcích.
b) Budeme-li předpokládat, že nevznikají chyby, tak úplné šetření poskytuje zcela přesné charakteristiky souboru (míry polohy, variability,

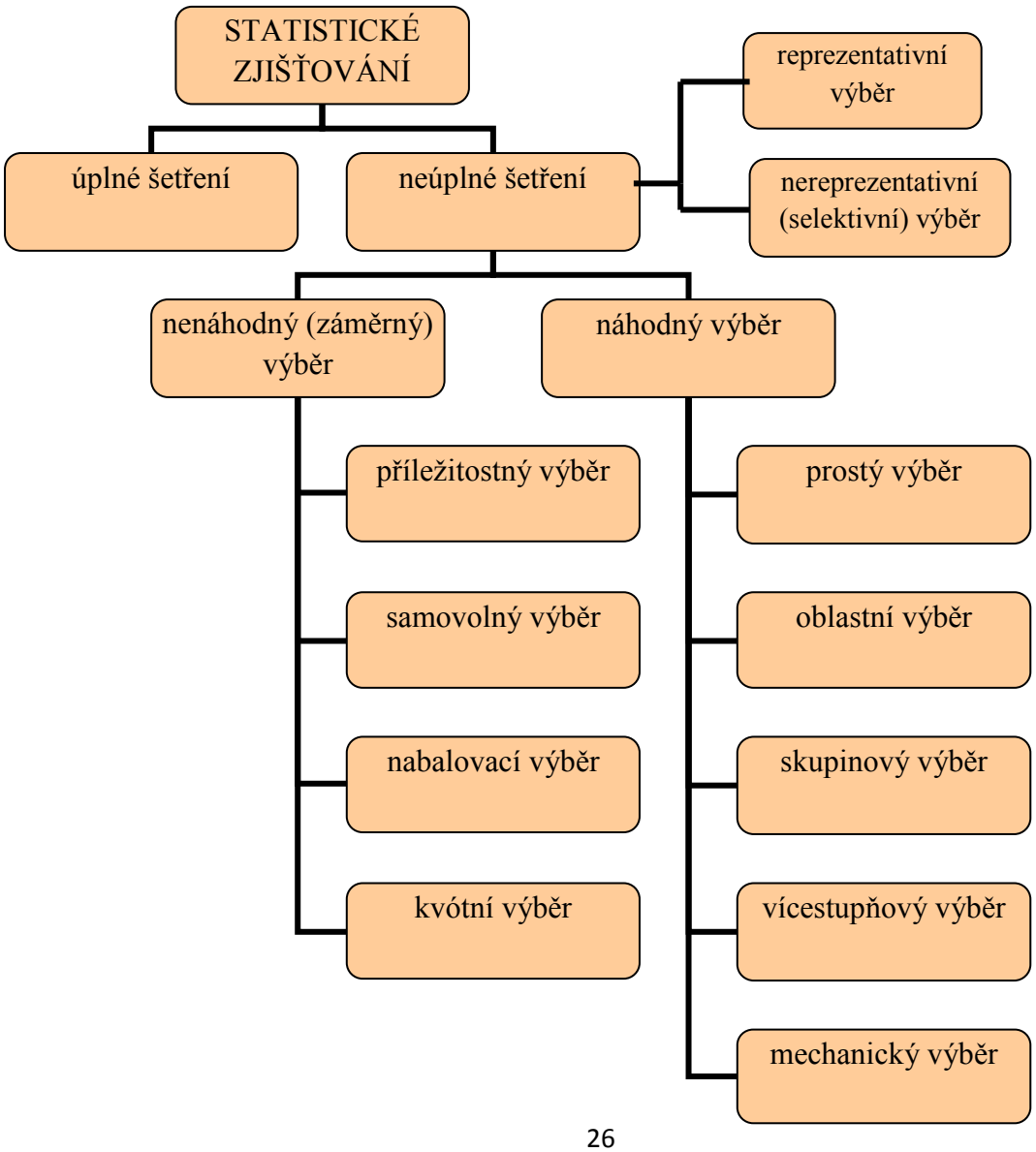
nezávislosti, atp.), zatímco neúplné šetření poskytuje pouze přibližné hodnoty (odhady) těchto charakteristik.

c) Úplná zjišťování se setkávají s větším pochopením u respondentů (osob, institucí) než šetření neúplná. [5]

5.2.1 VÝBĚROVÉ POSTUPY

Výběrový soubor je definován jako skupina statistických jednotek (= nositel sledované vlastnosti) ze základního souboru. Pro některé výběrové postupy je třeba *opora výběru*, což může být seznam, registr, databáze nebo mapa všech jednotek v souboru (například telefonní seznam). Jak tento výběrový soubor dostaneme? Zde je uvedeno několik výběrových metod.

Obrázek č. 4 Přehled výběrových postupů



Nereprezentativní (selektivní) výběr. Takový výběr poskytuje zkreslené informace o celkové populaci. Příkladem selektivního výběru může být například vzorek dětí vybraných z 9.třídy základní školy zaměřené na matematiku, pokud chceme zkoumat znalosti matematiky žáků 9. tříd základních škol.

Reprezentativní výběr. Tento výběr musí odrážet svým složením vzhledem ke sledovaným znakům vlastnosti celé populace.

Nepravděpodobnostní (záměrný) výběr. V tomto případě je subjektivně rozhodováno, které jednotky patří do výběru a které ne.

Pravděpodobnostní (také náhodný) výběr znamená, že každé jednotce ze souboru (populace) je přiřazena známá pravděpodobnost zahrnutí do výběrového souboru. Zde se občas setkáváme s pochybností, zda je tento výběr dobrým podkladem pro zabezpečení reprezentativnosti. Všem jednotkám předem přiřadíme známou pravděpodobnost – a to pravděpodobnost vybrání do vzorku.

Prostý náhodný výběr je výběr jednotek se stejnými pravděpodobnostmi vybrání (zahrnutí do vzorku - výběru). Zajišťuje stejnou pravděpodobnost zahrnutí do vzorku a rovněž všem vzorkům o stanoveném rozsahu stejnou pravděpodobnost jejich vzniku. Jedná se o nejjednodušší metodu pravděpodobnostního výběru. Tento výběr je prováděn různými technikami losování.

Oblastní výběr. Základní populace je rozdělena do dílčích skupin (oblastí). Skupiny jsou tvořeny tak, aby sledované znaky ve skupinách byly homogenní (příliš se nelišily) a skupiny mezi sebou byly zase naopak heterogenní (lišily se co nejvíce). Například při šetření na obyvatelstvu jsou oblasti tvořeny územními celky, věkovými skupinami apod. Je-li variabilita (proměnlivost) ve skupinách značně rozdílná, musí to být při rozvržení výběru zohledněno – nemá smysl vybírat mnoho jednotek z relativně stejnorodé a málo jednotek z malé relativně nestejnorodé skupiny.

Skupinový výběr. Tentokrát je soubor rozdělen do skupin, které jsou vnitřně různorodé, ale vzájemně se velmi podobají. A navíc skupiny jsou přibližně stejně velké. Požaduje se tedy, aby variabilita mezi skupinami byla co nejmenší. U obyvatelstva se nevybírají jednotlivé osoby, ale celé skupiny (rodina, škola, …).

Vícestupňový výběr. Je-li soubor příliš rozsáhlý, přímý výběr jednotek je prakticky neproveditelný, aniž bychom tento soubor roztřídili do menších skupin, často

v několika stupních. Na každém stupni je proveden náhodný výběr skupin. K posledním prvkům se dostáváme přes vyšší výběrové jednotky, např. města – bloky – domy – domácnosti.

Mechanický výběr. Tento výběr je založen na předem daném uspořádání prvků populace. Do výběru patří všechny prvky, které jsou od sebe vzdáleny o zvolený výběrový krok, přičemž první prvek je vybrán prostým náhodným výběrem.

Při sběru dat, která nezaručují reprezentativnost, se používají **příležitostné výběry** (například náhodné oslovení chodců na ulici) a **samovolné výběry** – *ankety*. Problém spočívá v tom, že populace, z níž jsou tyto výběry prováděny a na niž bychom mohli získané údaje zobecnit, je neznámá. Mohou ji tvořit lidé, kteří mají pozitivní vztah k anketám, v kladném (obdivovatelé) slova smyslu i záporném (stěžovatelé) slova smyslu. Nebo osoby, které jsou motivovány k účasti slibem výhry či odměny.

Nabalovací výběr. Zde se od již kontaktovaných jednotek získávají informace o dalších jednotkách, které by mohly patřit do zkoumané populace. Jedná se o nejrychlejší a nejlevnější typ výběru.

Kvótní výběry. Tento výběr je nejrozšířenějším výběrovým postupem. Shodu se základním souborem mají zajistit kvótní znaky stanovené expertem. Měly by být jednoduché, jednoduše identifikovatelné a z hlediska souboru klíčové vlastnosti, nejčastěji věk, pohlaví, vzdělání. Podle hodnot kvótních znaků je základní soubor rozdělen do skupin.

5.3 SBĚR DAT

K obstarání primárních dat je třeba hodně času i financí, jak již bylo uvedeno. Je to však nezbytné, zvláště tehdy potřebujeme-li aktuální informace. Ke snížení nákladů se používají sekundární data. Jsou to data, která jsou rychle k dispozici, jsou již známa. Je to sice levnější postup, ale data zase naopak nemusí být vždy aktuální.

Ve fázi sběru primárních dat si musíme odpovědět na několik otázek:

a) Co a kdo je statistickou jednotkou, jak je definován základní soubor a jaký je rozsah souboru?

b) Jaký bude způsob provedení výběru z toho souboru a jak bude velký rozsah výběru?
c) Jak budou zjišťovány a zaznamenány informace?
d) Jakých chyb se můžeme dopustit a jak se jich vyvarovat?

Základní kritérium pro volbu postupu sběru primárních dat plyne z charakteru zkoumaných jevů – zda jsou nebo nejsou přímo pozorovatelné, jsou-li spontánní, uměle vyvolané pro průzkum, zda jsou vázány na jednotlivce, skupiny osob apod. Primární data lze získat pozorováním, ale hlavně dotazováním. Jiné metody jsou obměnou těchto dvou základních.

Pozorování nám poskytuje objektivní a přesné údaje, neboť jsou získány přímo, nikoliv prostřednictvím jiné osoby. Musíme předpokládat, že jev, který má být pozorován, je jednoznačně vymezen a jeho projevy jsou zaznamenány. Nejjednodušším výsledkem pozorování je počet určitých jednotek (např. využití večerního tramvajového spoje) nebo fakt, že jev nastal či nikoliv. Pozorování složitějších jevů lze zjednodušit, jsou-li připraveny alespoň nějaké kategorie, v níž se může vyskytovat (např. reakce diváků na film – žádná, smích, pískot, potlesk či jiná). Pozorování může probíhat skrytě, ale i zjevně (pozorovatel je znám pozorovaným osobám). Zjevné pozorování ovšem může sledované skutečnosti ovlivnit a zásadně je změnit.

Dotazování. U této metody jsou údaje získávány prostřednictvím cílených otázek, buď pomocí rozhovoru tváří v tvář, telefonicky nebo písemnou formou – vyplňováním papírového či elektronického dotazníku. Obsah a forma rozhovoru či dotazníku může zásadně ovlivnit kvalitu získaných informací a možnosti jejich dalšího zpracování a využití. Firmy zabývajícími se průzkumy disponují s tzv. *tazatelskou sítí*. Jedná se o skupinu lidí, kteří jsou vyškoleni k tomu, aby prováděli vlastní kontaktování vybraných osob a zaznamenávali požadované informace.

5.3.1 TVORBA DOTAZNÍKŮ

Osobní dotazování je jednou z forem získání dat. Informace získané touto metodou jsou zapisovány do dotazníku. Vede-li rozhovor tazatel, vyplňuje odpovědi respondenta do dotazníku. Tato metoda zjišťování údajů zabraňuje nepochopení či špatnému pochopení

otázek, nečitelným údajům apod. Často se stává, že respondenta je těžké získat z důvodu narušení soukromí (např. dotazování v domácnostech). V tomto případě se tazatel s respondentem domluví na předání dotazníku a zároveň na termínu jeho vyzvednutí. Přímý kontakt s respondentem může také negativně ovlivnit důvěru v anonymitu průzkumu.

V dnešní době se velmi často využívá telefonních linek k dotazování, kdy se určitým způsobem zachovává osobní forma rozhovoru a méně zasahuje do soukromí. Tato metoda je velmi rychlá a umožňuje rychlé zpracování údajů. Na druhou stranu otázky musí být srozumitelné, nelze se zabývat detaily a rozhovor by neměl být příliš dlouhý.

Dotazník lze posílat i poštou. Výhodou této metody jsou nízké náklady, vyloučení vlivu tazatele a v neposlední řadě nižší náročnost organizace sběru dat. Největší nevýhodou je nízká návratnost dotazníků. Ke zvýšení návratnosti lze přispět opakovaným upozorňováním či prosbou, zda by mohl být dotazník vrácen apod.

Další způsob dotazování je pomocí elektronického dotazníku. Díky Internetu můžeme rychleji distribuovat dotazníky mezi respondenty. Na rozdíl od telefonních průzkumů má použití elektronického dotazníku svou nevýhodu - specifickou populaci, ne každý totiž používá Internet.

Forma dotazníku

Pro úspěšný výsledek dotazování jsou důležité i maličkosti, např. úprava, typ a velikost písma, barevnost, srozumitelné nákresy, schémata, atd. Základní pravidla pro tvorbu dotazníku z hlediska obsahu jsou tato:

- Bude respondent otázkám rozumět?
- Bude dotazovaný schopen na otázky odpovědět?
- Bude respondent ochoten odpovědět na otázky?

Kvalitu získaných informací ovlivňuje nejen způsob položení otázek, ale i pořadí, obsah a forma otázek.

Druhy otázek

Otázky podle pořadí. V úvodu by měly být použity formulace k objasnění účelu průzkumu, k navázání kontaktu s respondentem a získání jeho účasti v šetření – „*Bez Vaší*

pomoci se nám nepodaří ...“ apod. Respondent považující daný průzkum za prospěšný a své odpovědi za užitečné, bude s vyšší pravděpodobností ochoten odpovídat. Otázky by měly volně navazovat, aby neustále udržovaly soustředění a neodváděly pozornost přeskakováním z jednoho problému na jiný. Jestliže se v dotazníku vyskytuje více otázek se stejnou nabídkou možných odpovědí, je výhodné je spojit v tzv. *baterii otázek.*

Tabulka č. 2 Baterie otázek

Označte u každého časopisu, jak často jej kupujete	pravidelně	příležitostně	vůbec ne
Epocha			
21.století			
Lidé a země			
jiné			

Otázky podle účelu. Otázky zaměřené na podstatu dotazníku (meritorní otázky) mohou být zpestřeny užitím různých pomůcek, grafů či obrázků. *Pomocné* otázky slouží k vedení rozhovoru žádoucím způsobem. *Větvící* otázky slouží k rozdělení respondentů do určitých skupin, které dále odpovídají na odlišné otázky (například podle toho jak tráví svůj volný čas). *Filtrační* otázky identifikují osoby, které mohou poskytnout určitou informaci. Dotazník by měl být stručný a měl by se vyvarovat zbytečných otázek, které dotazník prodlužují. Zbytečné jsou otázky, které

- nesouvisí s cílem výzkumu či jsou zaměřeny na vedlejší či okrajové informace,
- přinesou jen velmi nespolehlivé odpovědi,
- mají jen nepatrnou variabilitu odpovědi,
- jsou v podstatě duplicitní (nadbytečné).

Abychom se vyvarovali nekvalitních odpovědí, používáme *kontrolní* otázky, kde mezi odpověďmi na tyto otázky a otázky položené dříve existuje souvislost. Kontrolní otázky mohou upozornit na špatně formulované a respondenty nesprávně pochopené

podstatné otázky, ale i na chyby v odpovědích. Tyto otázky v některých případech mohou odhalit tazatelské podvody, jako je doplňování údajů či vyplňování celých dotazníků.

Otázky podle obsahu. Otázky z hlediska obsahu dělíme na *přímé* a *nepřímé*. Přímé otázky slouží k vědomé odpovědi respondenta („Jakého jste vyznání?“). Tyto otázky nemusí být vždy příjemné, mohou vyvolat i pocity napětí. Tázaný může místo pravdivé odpovědi dát přednost „společensky přijatelnější“ odpovědi. Stejně tak nepříjemné mohou být otázky vyžadující určitou znalost („Souhlasíte s udělením Nobelovy ceny za matematiku?“), otázky na příčinu („Proč jste věřící?“) či provedení výpočtu („Kolik dělá čistý měsíční příjem Vaší domácnosti na hlavu?“). Zkreslení opovědí se dá vyvarovat zvolením lepší formulace přímé otázky, např. namísto „Jak často uklízíte koupelnu?“ zvolíme formulaci „Uklízíte koupelnu 2x týdně a méně často?“ nebo odvedeme pozornost od osoby tázaného, např. namísto „Jste nezaměstnaný?“ se zeptáme „Je ve Vaší domácnosti někdo nezaměstnaný?“ apod. Nepřímé otázky používáme tehdy, když se názor respondenta promítá do jiné osoby - přímá otázka: „Jste obětí domácího násilí?“ nepřímá otázka: „V dnešní době přibývá obětí domácího násilí, co na to říkáte?“ [2]

Otázky podle formy. Nejčastěji se používají *uzavřené* otázky. Tady respondent vybírá odpovědi z určitého okruhu dvou (*alternativní* otázky) či více možností (*selektivní* otázky). Jednotlivé možnosti odpovědí by měly pokrýt celý problém, neměly by se překrývat. Seznam odpovědí by neměl být příliš dlouhý, jinak se může stát, že respondent bude vybírat jen z několika prvních či posledních možností. Někdy můžeme tyto otázky obohatit o respondentovy odpovědi, potom mluvíme o *polouzavřené* otázce. U *otevřených* otázek respondent odpovídá sám. Jednoduše odpoví na otázky typu „Vlastníte domácího mazlíčka?“ nebo „Vlastníte jízdní kolo?“. Ne na všechny otázky je příjemné odpovědět, proto se využívá kategorizace, vytvoření intervalů. Místo otázky „Kolik je Váš čistý měsíční příjem?“ se zvolí např.

„Do které skupiny spadá Váš čistý příjem?“ (uvedeno v tisících Kč)			
<8 – 12 >	<13 – 18 >	<19 - 27>	<28 – a výše>

Kategorizováním se z otevřené otázky stává *uzavřená* otázka.

Nevhodná forma otázek komplikuje zpracování odpovědí, ale komplikuje i dotazování. Například otázkou „Máte automobil?“ není specifikováno, jestli vlastní auto

dotazovaný nebo někdo jiný z rodiny. U otázky „Co byste si pořídila, kdybyste vyhrála velkou sumu?“ zase není určeno, co je myšleno velkou sumou. Každý může mít (a určitě má) na mysli úplně jinou částku.

5.3.2 ZNAKY A JEJICH KVANTIFIKACE

Jednoduše změříme délku stolu, obvod pasu nebo zjistíme nejpoužívanější značku dámských šamponů. Mnohé informace získané od respondentů v průzkumech se však týkají stavu jejich mysli, jejich názorů a postojů. A právě měření těchto „veličin“ je v mnoha průzkumech velmi podstatné. „Měření“ mínění respondenta je velmi obtížné a ještě složitější je pro takové měření vybrat vhodnou stupnici (škálu). Představuje to převedení nejrůznějších znaků na stupně vybrané stupnice. Sama tvorba stupnice je velmi složitá a vyžaduje mnoho znalostí. V praxi se nejčastěji upřednostňují jednoduší a méně obtížné postupy.

Volba konkrétního postupu při práci s daty závisí na charakteru dat. Ne pro každou proměnou platí, že lze mluvit o její vyšší či nižší hodnotě nebo že nula vyjadřuje naprostou absenci znaku (např. teplota ve stupních Celsia, 0°C neznamená, že žádná teplota není; zatímco 0°K znamená, že nelze už odebírat z konkrétní látky žádné teplo = absolutní nula). Díky použitému způsobu vyjadřování hodnot sledované veličiny rozlišujeme kvantitativní a kvalitativní znaky, resp. nominální, ordinální, kardinální proměnné (znaky), intervalové znaky, atd.

Kvantitativní znaky se vyjadřují pouze čísly, obvykle charakterizují množství, velikost. **Kvalitativní znaky** vyjadřují určitou vlastnost a jsou vyjádřeny slovně, např. zaměstnání obyvatel. [11]

Mezi hodnotami **nominálních** (jmenných) proměnných se uplatňuje vztah ekvivalence – neekvivalence. Neexistuje objektivní kritérium pro uspořádání jejich hodnot (kategorií) „podle velikostí“. Hodnoty nominálních proměnných jsou obvykle vyjadřovány slovem, čísla se používají jako kódy a mohou být libovolně volena.

Ordinální (pořadové) proměnné vyjadřují vyšší či nižší stupeň sledovaného znaku. Kategorie lze objektivně uspořádat od nejnižšího po nevyšší stupeň, ale čísla pouze reprezentují pořadí, nikoliv kvantitu ve smyslu počtu měrných jednotek (pořadí lze vyjádřit čísly 1, 2, 3, stejně tak čísly 10, 12, 22).

Hodnota **kardinální** (měřitelné, poměrové) proměnné představuje určitý počet konstantních měrných jednotek, které se používají pro vyjádření stupně sledovaného znaku nebo vyjadřuje počet nějakých objektů či osob. Nabývají pouze kladných hodnot jejich obměny lze porovnávat jak rozdílem, tak podílem. Je tedy možno měřit o kolik měrných jednotek je jedna obměna větší (menší) než druhá a také kolikrát je jedna obměna větší (menší) než druhá. Kardinální intervalové stupnice se příliš nepoužívají. Velmi často se používají jednoduché pořadové škály (známkovací, bodovací), jejichž hodnoty jsou vyjadřovány přirozenými čísly. Hodnoty na stupnici jsou vnímány jako stejně vzdáleny, i když tomu tak není (např. vzdálenost známek 2 a 1 nemusí být stejná jako vzdálenost známek 3 a 2).

Intervalové znaky. Zde se předpokládá, že mezi sousedními hodnotami jsou stejné vzdálenosti.

Dalším dělením, pouze u kvantitativních znaků, je dělení na spojité a diskrétní znaky. **Diskrétní znaky** nabývají konečného počtu či spočteného počtu hodnot, například počet zaměstnanců Gymnázia Jana Opletala v Litovli. **Spojité znaky** mohou nabývat nekonečně mnoho hodnot z konečného nebo nekonečného intervalu.

Dalším krokem je vytvoření měřící stupnice. Samotné vytvoření stupnice vyžaduje aplikace pracných škálovacích postupů. Jejich cílem je vytvořit na základě skupiny neměřitelných znaků vyjadřujících sledovaný jev jednorozměrnou stupnici hodnot, která slouží k „měření“ sledovaného jevu.

V průzkumech se nejčastěji používají bodovací (známkovací) stupnice. Tvoří je jednoznačné uspořádané hodnoty, číselné (1, 2, 3, 4, 5) nebo slovní (výborně, velmi dobře, dobře, dostatečně, nedostatečně). Tato stupnice je oblíbená pro její jednoduché použití. Ale i tady se musí dodržovat několik pravidel. V první řadě je nutno stanovit počet stupnicových hodnot. Zpravidla je lepší volit lichý počet hodnot, neboť umožňuje umístění do středu stupnice „neutrální bod“. Počet stupnicových hodnot před a za tímto bodem se shoduje.

Další možností jak lze vyjádřit odstupňování sledovaného jevu je grafické vyjádření. Škálové stupně můžou představovat například nádoby od prázdné, postupně zaplněné až po plnou, apod. [2] Stupeň sledovaného znaku lze graficky vyjádřit jako umístění bodu na úsečku, kde krajní body úsečky představují opačné extrémy.

Jste spokojeni s výukou cizích jazyků na naší katedře?

Zcela spokojen Zcela nespokojen

Jestliže v daném souboru zkoumáme výskyt jednoho znaku, mluvíme o *jednorozměrném statistickém souboru*. Výsledkem třídění souboru podle jedné veličiny je tabulka nebo graf. Příkladem takovéhoto jednorozměrného souboru je soubor studentů, u kterých se zjišťuje počet mimoškolních aktivit. Prvním krokem je uspořádat data získaná určitým šetřením. Lze-li při třídění dat jednoznačně uspořádat hodnoty znaku podle velikosti, přiřadit jim příslušné četnosti (absolutní nebo relativní), pak tabulka nebo graf zobrazuje tzv. bodové rozdělení četností (u diskrétních znaků s malým počtem variant – obměn nebo ordinálních či nominálních znaků).

Tabulka č. 3 Tabulka rozdělení četností

a_j	a_1	…	a_k
n_j	n_1	…	n_k

Kde a_j je hodnota znaku a n_j je kolikrát se a_j vyskytla v souboru.

Absolutní četnost je definována jako počet výskytů jednotlivých variant znaku, obvykle se značí n_1, n_2, …, n_k. Rozsah soubor se značí n. Vypočítá se jako součet četností všech variant, tj.

$$\sum_{j=1}^{k} n_j = n \quad \text{j =1,2,3,…,k}$$

Relativní četnost je poměr absolutní četnosti dané varianty vůči rozsahu souboru. Většinou se uvádí v procentech. Vypočítá se pomocí vztahu

$$p_j = \frac{n_j}{n},$$

přičemž platí

$$\sum_{j=1}^{k} p_j = \sum_{j=1}^{k} \frac{n_j}{n} = \frac{1}{n}\sum_{j=1}^{k} n_j = 1$$

Dále je definována *kumulativní četnost*, kterou získáme postupným přičítáním absolutních nebo relativních četností jednotlivých variant znaku. Pokud je zkoumán spojitý statistický znak nebo diskrétní znak s mnoho variantami, potom počítáme, tzv. intervalové četnosti. Intervaly (třídy) se obvykle volí stejně dlouhé – ekvidistantní. V každém intervalu je zvoleno číslo, které jej zastupuje – střed nebo reprezentant intervalu (třídy). Tak dostáváme tabulku intervalového rozdělení četností (a_j, j=1,…k, je střed intervalu, c_i, i=0, …, k, krajní body – hranice třídy a n_j, j=1, …, k, intervalová četnost).

Tabulka č.4 Intervalové rozdělení četností

interval	$(c_0, c_1>$	$(c_1, c_2>$	…	$(c_{k-2}, c_{k-1}>$	$(c_{k-1}, c_k>$
a_j	a_1	a_2	…	a_{k-1}	a_k
n_j	n_1	n_2	…	n_{k-1}	n_k

Ve statistických souborech často potřebujeme určit hodnotu, kolem které se data soustřeďují, stanoví se jejich „střed“. Těmto číslům, říkáme míry (charakteristiky) polohy.
Aritmetický průměr :

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{k} a_j n_j \ .$$

Geometrický průměr se používá právě tehdy, jsou-li hodnoty znaku kladné. Vzorec:

$$\bar{x}_G = \sqrt[n]{(a_1)^{n_1}(a_2)^{n_2}\ldots(a_k)^{n_k}} \ .$$

Harmonický průměr se používá při charakterizaci úrovně znaku, jehož hodnoty lze vyjádřit jako poměr dvou jiných hodnot. Též se používá jen při kladných hodnotách znaku. [12]

$$\bar{x}_H = \frac{n}{\sum_{1}^{k}\frac{n_j}{a_j}}.$$

Dané vzorečky se používají, jsou-li data uspořádaná do četnostní tabulky.

Dalšími charakteristikami jsou modus a medián. *Medián* $\tilde{x}$ je „prostřední“ hodnota statistického znaku, která dělí soubor na dvě stejně velké skupiny. Velikost mediánu určíme na základě pořadí hodnot zkoumaného statistického znaku v uspořádaném souboru. Nejprve se tedy statistický soubor musí seřadit podle velikosti hodnot znaku.

Pro liché n pak platí: $\tilde{x} = x_{\frac{n+1}{2}}$.

Pro sudé n platí: $\tilde{x} = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$.

Modus je nejčastěji se vyskytující hodnota znaku v daném statistickém souboru.

Při třídění souboru dat podle dvou znaků (*dvourozměrný statistický soubor*) můžeme data uspořádat do tzv. kontingenční tabulky, která obsahuje četnosti všech možných dvojic variant sledovaných variant znaků X, Y

Tabulka č. 5 Kontingenční tabulka

X\Y	1	2	…	s	∑
1	n_{11}	n_{12}	…	n_{1s}	$n_{1.}$
2	n_{21}	n_{22}	…	n_{2s}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
r	n_{r1}	n_{r2}	…	n_{rs}	$n_{r.}$
∑	$n_{.1}$	$n_{.2}$	…	$n_{.s}$	n

n_{ij} je četnost jevu [X = i, Y = j], pozorovaná (empirická) četnost

$n_{i.}$, $n_{.j}$ jsou marginální četnosti, nacházejí se na okrajích tabulky

$$\sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{s} n_{.j} = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = n, \qquad n_{i.} = \sum_{j=1}^{s} n_{ij}, \quad n_{.j} = \sum_{i=1}^{r} n_{ij}$$

V některých případech nás zajímá, zdali znak Y závisí na znaku X případně naopak. O této závislosti dokážeme rozhodnout právě na základě analýzy kontingenční tabulky (např. χ^2 testem nezávislosti). Musíme pamatovat na to, že statistickými metodami nejsme schopni rozhodnout, zda zkoumaný vztah je příčinný (kauzální) nebo zda nejde o závislost pouze zdánlivou.

5.3.3 CHYBY PŘI SBĚRU DAT

V průběhu průzkumu vznikají chyby dvojího druhu: *náhodné chyby* - ovlivňují přesnost prováděných závěrů o populaci, a *systematické chyby* – zkreslují závěry o populaci. [2] V rámci náhodných chyb se vyskytují *náhodné výběrové chyby*. Ty vznikají tehdy, jestliže prošetření jednoho výběrového souboru přinese jiné výsledky, než šetření u jiného výběrového souboru. Můžeme je omezit, zvýšíme-li rozsah výběrového souboru a tím dosáhneme i nižší variability výběrové statistiky. Zvýšením rozsahu výběru se zpřesní odhady populačních charakteristik – parametrů.

Při průzkumech, které vznikají dotazováním obyvatelstva, může dojít k systematickým chybám (odchylkám). Tyto chyby se mohou vyskytnout ve všech fázích sběru dat. Jejich zdroje tkví ve způsobu zjišťování údajů. Nenáhodné chyby mohou být vyvolány několika způsoby [2]:

a) Systematické zkreslení může být vyvoláno nesprávnou představou o populaci. Hlavně v těch případech, kdy máme různé typy samovolných výběrů, může být prováděno zobecňování a interpretace výsledků pro jinou populaci, než je tomu ve skutečnosti.

b) Máme-li správně vymezenou populaci, může systematické zkreslení vzniknout i nereprezentativním výběrem. Nereprezentativní výběr znamená, že vzorek nepokryje určitou cílovou populaci nebo naopak nadměrně zastupuje jinou část populace. Minimalizovat tuto chybu můžeme vhodně zvoleným časem, místem i způsobem zjišťování, výběrem a přípravou tazatelů, atd.

c) Zkreslení může nastat *na straně respondenta* (nechce či neumí odpovědět na otázku, nepochopil otázku, snaží se ukázat v lepším světle, apod.) i *tazatele* (jeho chování k dotazovanému, postoje, způsob vedení či záznam dotazování). Zdrojem zkreslení může být i délka dotazníku či rozhovoru, nezajímavý obsah, úprava dotazníku, atd.

Systematické chyby mají tendenci růst při zvětšujícím se rozsahu výběru, na rozdíl od náhodných chyb, které se zmenšují se zvyšujícím se počtem statistických jednotek výběru. Vzniku náhodné chyby zabránit nelze, avšak vhodnou volbou rozsahu výběrového souboru lze omezit její velikost na únosnou mez.

5.3.4 VYUŽITÍ POČÍTAČOVÉ TECHNIKY

V dnešní době se při zpracování dat hojně využívají počítače. Použití počítače má však i své nevýhody.

Výhody:

- **Přesnost a rychlost**. Dnešní softwary nám poskytují velmi rychle požadované výsledky. Při ručním zpracování dat se velmi často objevovaly aritmetické chyby a časově byly výpočty velmi náročné.
- **Univerzálnost**. Počítače zpřístupňují širokou škálu statistických metod a umožňují provést velmi rychle i rozsáhlé komplexní statistické analýzy.
- **Flexibilita.** Počítač umožňuje rychle provést nové zpracování i při rozsáhlejších změnách dat.
- **Velikost datového souboru.** Počítač slouží ke zpracování rozsáhlých souborů dat, jejichž ruční zpracování by bylo téměř nemyslitelné.
- **Snadný přenos dat**. Jakmile jsou data jednou v počítači, můžeme je poslat na e-mail, zaznamenat na CD či na jiné médium a dále s nimi manipulovat nebo je uchovat pro pozdější rozsáhlejší analýzy.

Nevýhody:

- **Chyby v softwaru.** Ne každý software je na 100% spolehlivý. Některé programy mohou poskytovat chybné výsledky, protože programátor udělal chybu při tvorbě programu nebo neporozuměl statistické metodě. Stejně tak se může stát, že uživatel není dostatečně srozuměn s funkcí softwaru, např. díky nedostatečné nápovědě či manuálu k programu a používá ho nesprávně.
- **Univerzálnost**. Tato vlastnost patří mezi výhody, ale může být současně i nevýhodou. Existuje mnoho statistických metod pro zpracování dat, a tak se může snadno stát, že bude vybrána špatná metoda pro konkrétní zpracování dat.
- **Špatná data plodí špatné závěry**. Jestliže jsou data nasbírána špatně, nemůžeme předpokládat, že závěry z takových dat budou správné. Dále mohou být data poškozena tím, že se špatně zpracovávají datové soubory -

některé údaje mohou chybět, pokud jsou hodnoty chybně vloženy do počítače nebo se vyskytly nedostatky již při samotném sběru dat. [7]

5.4 PREZENTACE A INTERPRETACE DAT

Prezentace výsledků statistického šetření může být dána slovním výkladem, tabulkami i grafickým znázorněním.

Tabulky

V tabulkách jsou zaznamenány výsledky šetření v přehledném a srozumitelném tvaru. Při tvorbě tabulky by se měly dodržet určité zásady, které zmenší riziko nedorozumění při její interpretaci. Základními prvky tabulky jsou záhlaví tabulky, legenda tabulky, vlastní obsah tvořený řádky a sloupci tabulky. Součástí tabulky je i její název (stručný a výstižný) a může být uvedeno i číslo tabulky (je-li v dokumentu více tabulek). Používají se i poznámky, komentáře nebo vysvětlivky. Pro důvěryhodnost údajů uvedených v tabulce se uvádí pramen. Následuje příklad takovéto tabulky.

Tabulka č. 6 Přehled počtu požárů ve vybraných městech Olomouckého kraje v letech 2006-2009.

Přehled počtu požárů ve vybraných městech Olomouckého kraje v letech 2006-2009				
Název města	Počet zaznamenaných událostí			
	2006	2007	2008	2009
Litovel	40	34	42	34
Olomouc	325	338	320	323
Uničov	29	52	33	35
Prostějov	158	150	189	183
Mohelnice	36	38	39	30
Zábřeh	30	42	36	30
Jeseník	68	58	69	77

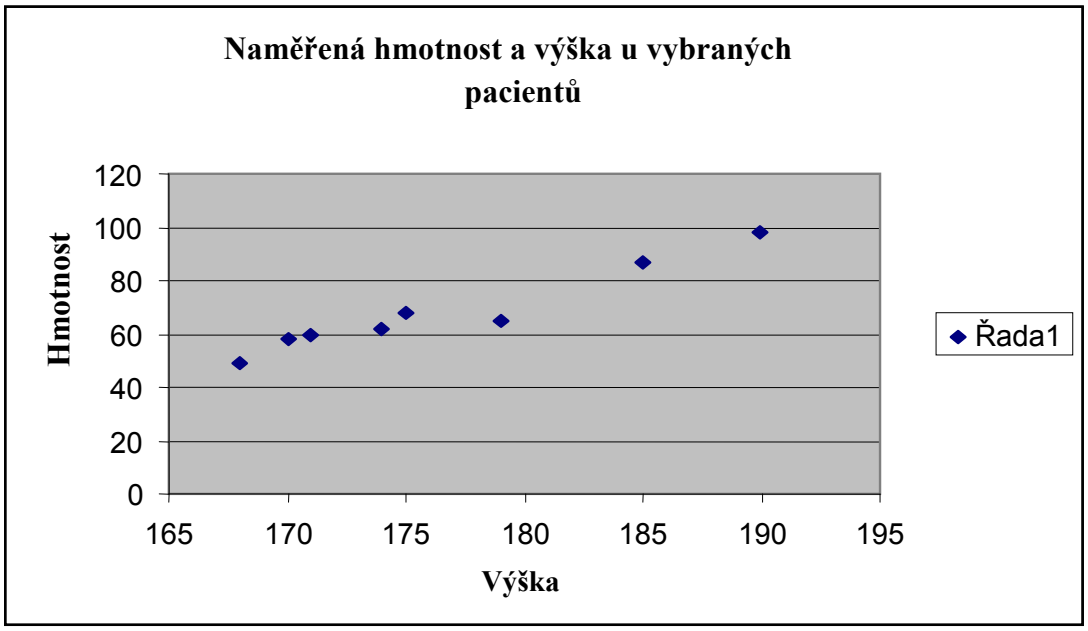
Zdroj: Hasičsky záchranný sbor Olomouckého kraje[15]

Grafy

Grafické vyjádření je obecně pochopitelnější a názornější než tabulkové vyjádření. Grafické prostředky používané při vytváření grafů jsou úsečky, body, křivky, plošné geometrické obrazce, tělesa části map, atd. Grafy jsou vyjadřovány buď v rovině (tzv. 2D grafy) nebo v prostoru (3D grafy). Podobně jako u tabulky je součástí grafu nadpis, číslo, legenda (vysvětlující použití barev, čar, šrafování), poznámky a vysvětlivky. V dnešní době se grafy zpracovávají pomocí různých softwarů, „ruční“ kreslení téměř vymizelo. Pro prezentaci nasbíraných dat je používáno mnoho různých grafů, zde je uvedeno několik příkladů.

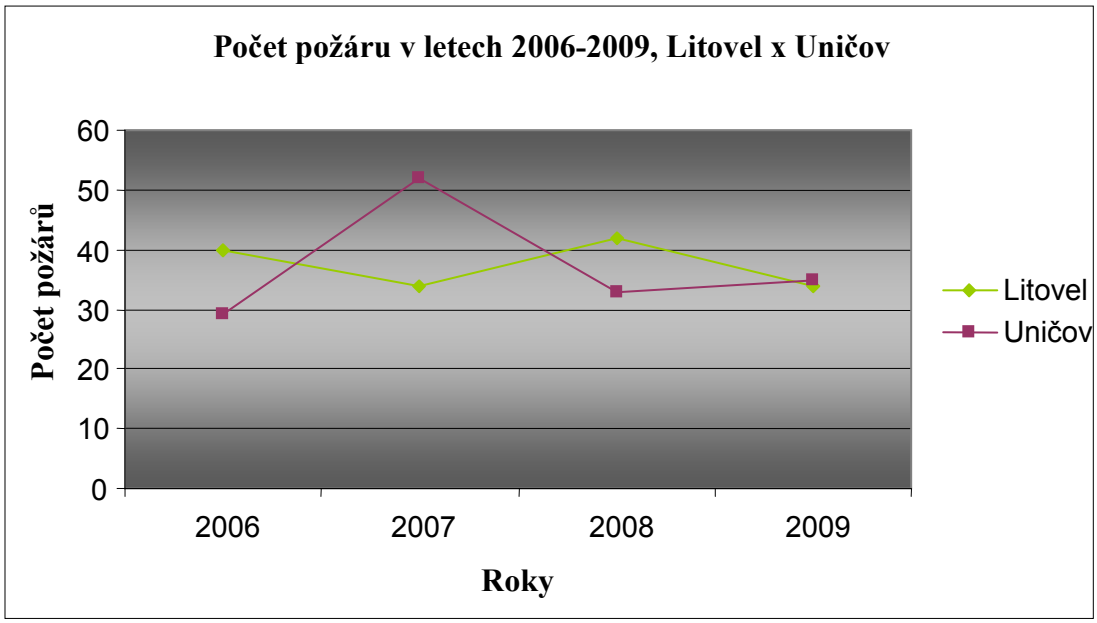
Bodový graf - v tomto grafu jsou znázorněny hodnoty v soustavě pravoúhlých souřadnic. Tento typ grafu se používá zejména tehdy, chceme-li vyjádřit závislost nějakého znaku na jiném znaku.

Obrázek č. 5

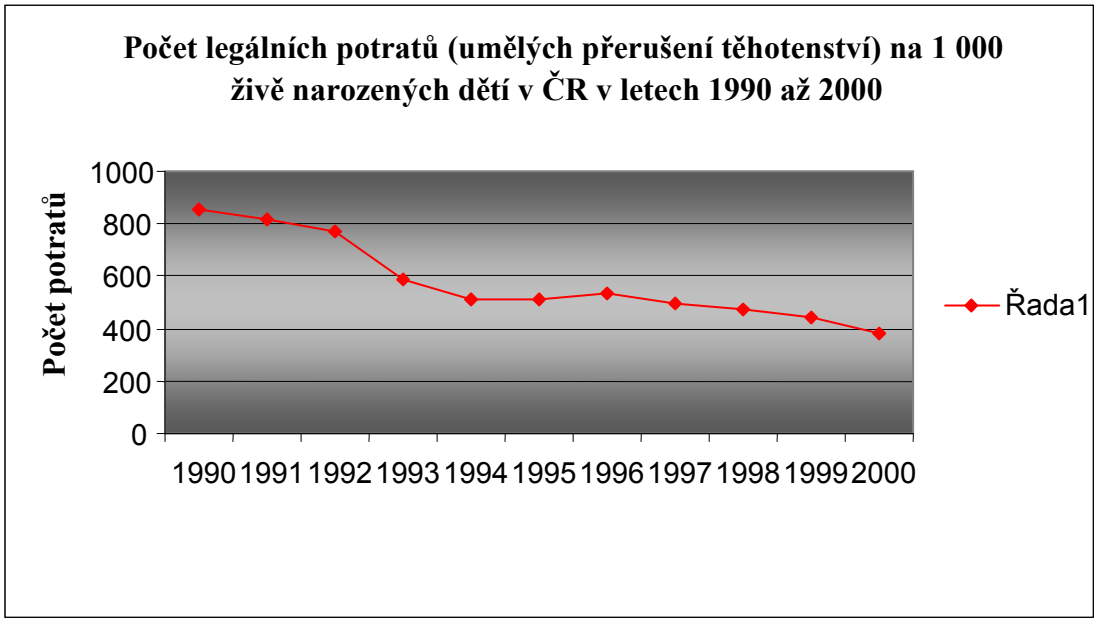


Spojnicový graf – grafickým prostředkem je čára, která vznikne spojením bodů. Hodnoty různých kategorií v jednom spojnicovém grafu odlišíme rozdílnými typy čar nebo rozdílnými barvami. Tento graf se používá ke znázornění průběhu časové řady (hlavně okamžikové) = časovému srovnání.

Obrázek č. 6

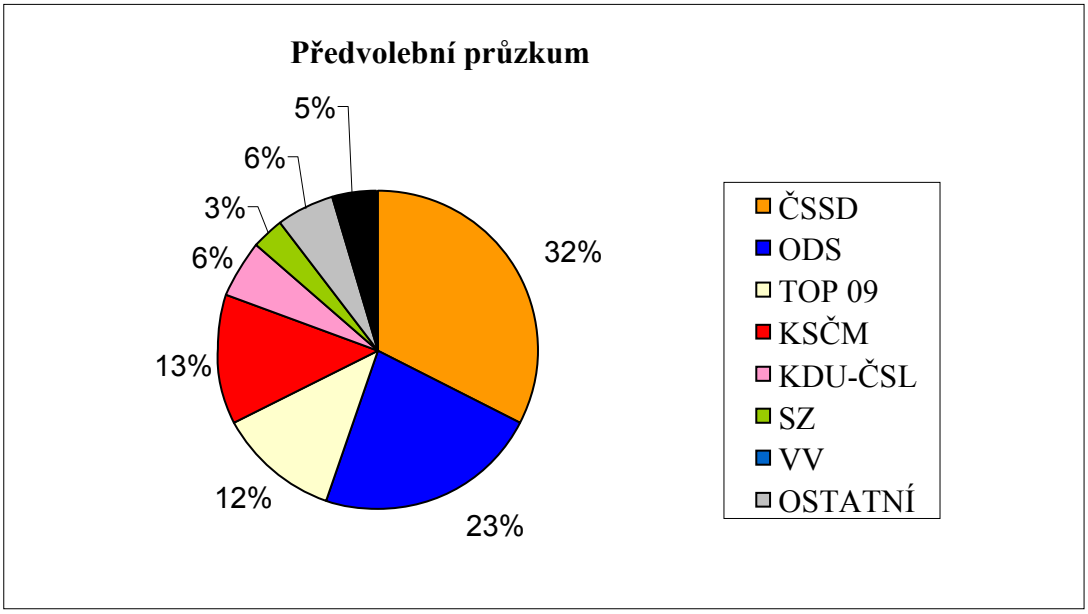


Obrázek č. 7



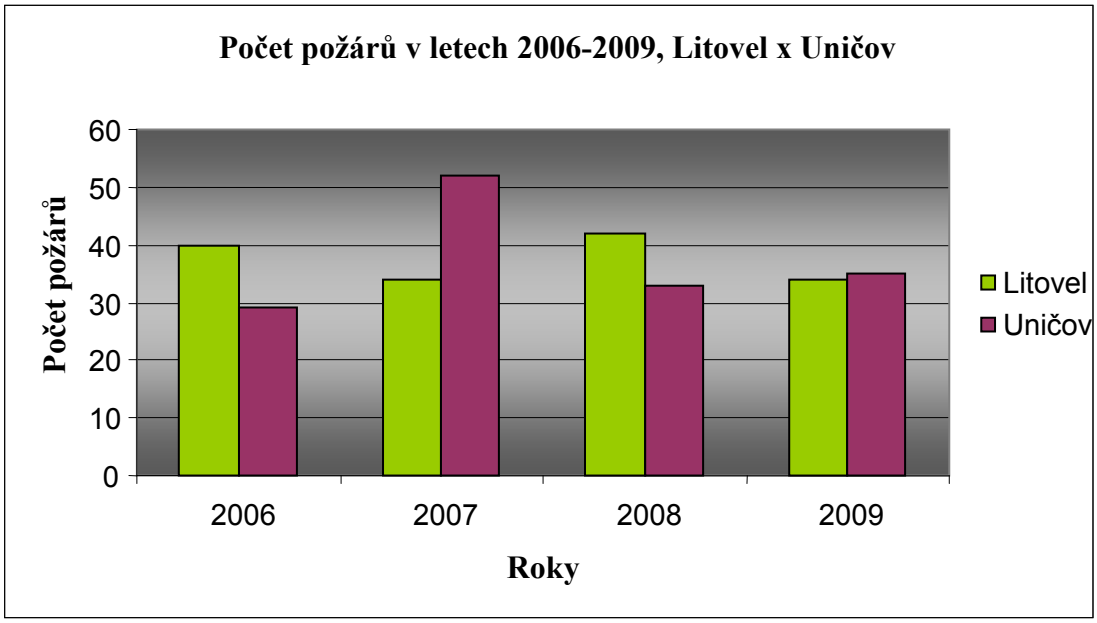
Výsečový graf – je speciální případ plošného grafu. Kruhové výseče, které představují jednotlivé části celku, odlišujeme různým šrafováním nebo barevně.

Obrázek č. 8



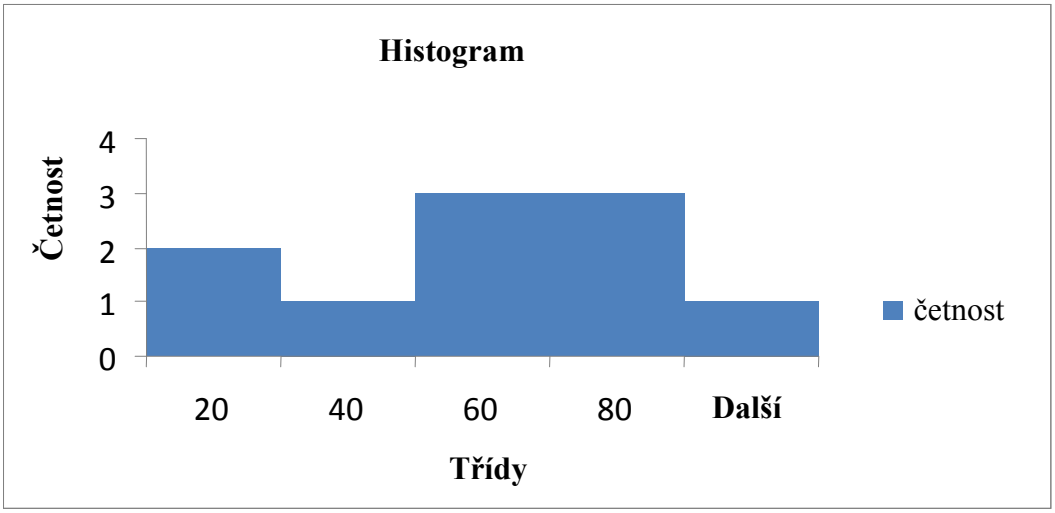
Sloupcový graf – v tomto grafu jsou číselné hodnoty vyjádřeny pomocí obdélníkových sloupců. Sloupce v grafu jsou obvykle zakreslovány svisle, v případě dlouhého textu u sloupců se zakreslují ve vodorovné poloze. Je-li v jednom znaku srovnáváno více souborů, umístí se do téže třídy i více sloupců. Sloupce se odlišují barevně nebo různým šrafováním.

Obrázek č. 9



Histogram – se používá ke znázornění intervalového rozdělení absolutních a relativních četností. Sloupce v histogramu jsou vždy vertikální, výška většinou odpovídá četnosti.

Obrázek č. 10



Tady je příklad jak by graf neměl vypadat. Graf je vytvořen tak, aby vzbuzoval dojem dramatického nárůstu nezaměstnanosti, i když tomu tak ve skutečnosti není. Osa y není popsána, i když jsou naznačeny záchytné body. Skoky mezi jednotlivými měsíci jsou neúměrné. [8]

Obrázek č. 11

Závěr

Cílem mé práce bylo upozornit na špatné statistiky či špatně interpretované statistiky. Každý den jsme zahlceni novými statistikami, zvláště nyní, v předvolebním období. Není média, ve kterém by nebyly uvedené grafy, tabulky, nebo které by se neodkazovalo na různé statistiky.

Tvůrci statistik si mohou s naším názorem „pohrát“, jak se jim zlíbí. Jak jste se z mé práce mohli dozvědět, záleží na mnoha faktorech při vytváření statistik: na konkrétní populaci, správném výběru, atd. Musíme se naučit číst ve statistikách. Jakmile uvidíte nebo uslyšíte novou statistiku musíte se ptát na několik základních otázek. To vám pomůže při vytváření názoru na daný problém. Nesmíme brát každé vyřčené číslo za pravdivé, odrážející skutečnou realitu. Musíme pátrat po vzniku uvedeného čísla.

Při psaní práce jsem se snažila, aby byla přínosem pro čtenáře, kteří se zajímají o vnitřní stránku statistik, které vidí každý den.

Použité zdroje:

[1] BEST, Joel *Dammend Lies*. Los Angeles : The Regents of the University of California, 2001. 190 s.

[2] PECÁKOVÁ, Iva. *Statistika v terénních průzkumech*. Praha 4 : Professional Publishing, 2008. 231 s. ISBN 978-80-86946-74-0.

[3] MOKRÝ, Radek . Pětina Británie je funkčně negramotná, v ČR je to 16 procent lidí. *Britské listy* [online]. 18.5.2004, [cit. 2010-03-12]. Dostupný z WWW: <http://www.blisty.cz/art/18171.html>.

[4] NOVOTNÝ, Radovan. Jak výběr respondentů při analýzách zkresluje závěry? . *Editorial* [online]. 27.7.2009, [cit. 2010-03-09]. Dostupný z WWW: <http://investujeme.cz/clanky/jak-vyber-respondentu-pri-analyzach-zkresluje-zavery/>.

[5] ŘEZANKOVÁ, Hana; MAREK, Luboš; VRABEC, Michal. *INTERAKTIVNÍ UČEBNICE STATISTIKY* [online]. Praha, 2000 [cit. 2010-02-21]. Úplná a neúplná statistická šetření, s. . Dostupné z WWW: <http://iastat.vse.cz/setreni1.html>.

[6] ZVÁROVÁ, Jana Úvod do statistické metodologie. In *Biomedicínská statistika*. Praha : EuroMISE UK, 2003 [cit. 2010-02-02]. Dostupné z WWW: <http://new.euromise.org/czech/tajne/ucebnice/html/html/node3.html>.

[7] ZVÁROVÁ, Jana Úvod do statistické metodologie. In *Biomedicínská statistika*. Praha : EuroMISE UK, 2003 [cit. 2010-02-11]. Dostupné z WWW: < http://new.euromise.org/czech/tajne/ucebnice/html/html/node4.html >.

[8] MINAŘÍK, Bohumil. *STATISTIKA I pro ekonomy a manažery*. Brno : Mendelovova zemědělská a lesnická univerzita v Brně, 1995. 158 s. ISBN 80-7157-166-0.

[9] *Http://www.hf.tul.cz/* [online]. [cit. 2010-01-12]. JEDNOROZMĚRNÁ POPISNÁ STATISTIKA. Dostupné z WWW: <http://www.hf.tul.cz/upload/files/mo-popis.pdf>.

[10] *Http://www.zf.jcu.cz/* [online]. 2005 [cit. 2010-02-18]. Aplikované kvantitativní metody pro zemědělskou praxi. Dostupné z WWW: <http://www2.zf.jcu.cz/public/departments/kmi/MSMT_05/zakladni%20pojmy.pdf>.

[11] - STRÁDALOVÁ, Jarmila ; KUBÁTOVÁ, Květa. *Vybrané kapitoly ze statistiky I.*. Praha 1 : Karolinum - nakladatelství Univerzity Karlovy, 1997. 250 s.

[12] KUNDEROVÁ, Pavla. *Základy pravděpodobnosti a matematické statistiky*. Olomouc, 2004. 186 s. ISBN 80-244-0813-9.

[13] JANEČEK, Arnošt; CHOVANEC, Kamil. *Http://www.median.cz/index.php?lang=cs* [online]. 2010 [cit. 2010-01-11]. Volební model stranické preference, vývoj voličských preferencí. Dostupné z WWW: <http://www.median.cz/docs/preference_2010_02.pdf>.

[14] *Http://www.volebni-preference.cz/* [online]. 2010 [cit. 2010-02-21]. Aktuální volební preference. Dostupné z WWW: <http://www.volebni-preference.cz/>.

[15] *Http://www.hzscr.cz/* [online], dostupné z WWW: http://www.hzsol.cz/statistika/rok-2006/

Seznam obrázků

Seznam tabulek