VPL 590 Základy výzkumu

(jarní semestr 2017)

Sekundární analýza dat

Ondřej Hora

Sekundární analýza dat

Práce s daty vytvořenými někým jiným anebo za jiným účelem.

Praktický význam a využitelnost

- u mnoha lidí jedna z nejčastějších situací v pracovním životě
- je důležitá, i když přímo neděláme sociálně vědní výzkum
- potřebnost dat je vysoká pro sociální pracovníky, úředníky státní zprávy a samosprávy
- jedná se o základní kompetenci

Zahrnuje např.:

- vyhledávání dat: potřebujeme vyhledat data
- využívání dat: abychom podpořili naše myšlenky (projektová žádost). Když děláme „velký výzkum“, může být sekundární analýza dat velkou výhodou
- vytváření dat: vytváříme data pro autority, musíme psát hlášení o počtu klientů, jejich struktuře atd.

Co jsou to data?

- Konkrétní údaje = popisují např. nějaký stav (věk 35 let), vlastnost (odpovědnost) nebo názor (legitimita sociální politiky).
- V kvantitativním výzkumu jsou zpravidla ve formě čísel nebo symbolů (slov) = existují různé jazyky dat
- Data získáváme během tzv. pozorování (observations) – proces mentální konstrukce dat (např. rozhovor s tazatelem).
- Data jsou zaznamenaná (empirická evidence = doklad) – např. do dotazníku, poznámkového archu či přímo do databáze (administrativní povahy).
- Data jsou vytvořena (manufactured) z hlediska výzkumu záměrně či nezáměrně (např. matriky).
- Data sbíráme a analyzujeme (podklad), abychom získali informace a znalosti (odpověděli na výzkumné otázky)
- Data sama o sobě jsou zpravidla neutrální (nenesou význam, ten jim přikládáme význam v procesu porozumění + konstrukce + hodnoty).

Cuesta (2013): problém, příprava dat, průzkum dat, modelování, vizualizace.

Analýza dat jako proces

kdy (jsou potřeba specifické dovednosti):

- Definujeme problém (vymezujeme, konceptualizujeme)
- Hledáme, získáváme, shromažďujeme data
- Snažíme se porozumět povaze dat (např. sledujeme meta-data tj. textové popisy o základních charakteristikách dat)
- Zvažujeme a třídíme data (posuzujeme jejich vhodnost z hlediska cíle) a vybíráme data
- Pokud je to potřeba, posuzujeme data z metodologického hlediska (např. jaká byla metoda výběru, musíme-li v datech uplatnit váhu)
- Provádíme průzkum dat (co je v datech, může být skryto…)
- Upravujeme data (čistíme data, redukujeme data, řešíme chybějící hodnoty, rekódujeme data …)
- Testujeme konstruktovou validitu použitých konceptů
- Provádíme případné statistické operace (např. regresní model)
- Prezentujeme data, připravujeme výstupy (vhodná forma)
- Prezentujeme určité významy prezentace konkrétních dat

Rozlišení primární a sekundární analýzy dat

Primární analýza dat

Je analýza dat získaných v našem výzkumu, kdy jsme provedli všechny předchozí potřebné kroky (stanovili-jsme si hlavní výzkumnou otázku, konceptualizovali pojmy, naplánovali výzkumný design a provedli šetření)

Sekundární analýza dat (analýza sekundárních dat)

Zahrnuje případy, kdy zpravidla jsem se neúčastnil žádné fáze výzkumu: tvorba dotazníku, sběr dat ani analýza (Michael a Eaton 2012).

- empiricismus a positivismus
- sjednocování metodologie (SA jako výzkumná technika)
- rozvoj počítačů a databázových systémů (umožnil skladování dat)
- zájem o mezinárodní výzkum + mnohé analýzy nelze v praxi provést jinak než sekundární analýzou (např. mezinárodní komparace LFS)
- navzdory zdánlivé snadnosti obsahuje úskalí a činní problémy
- Sekundární analýza dat je „dva kroky od reality“ neboť nemá kontrolu nad konceptualizací výzkumu ani nad sběrem dat (Blaikie 2003).

Možnosti sekundární analýzy dat:

1) dosud neanalyzovaná data (např. statistiky) jsou využita pro analýzu

- Běžně dostupná i „odložená“ data.
- Výzkumník pracuje i s daty, která v původním výzkumu mohla podlehnout redukci (část dotazníku nebyla pro analýzu použita).
- Ta samá data mohou být v jedné analýze primárními daty a v druhé sekundárními daty.

2) již využitá data jsou:

- analyzována někým jiným
- za jiným účelem (v jiném kontextu)
- jinou vhodnější metodou
- jde o ověření důležitých či překvapivých výsledků.
- jde o shrnutí poznatků výzkumu

Přínosy sekundární analýzy dat

- Relativní rychlost a finanční nenáročnost výzkumu (někdy stačí počítač s internetem nebo přístup do knihovny).
- Časová úspora (data již někdo sebral, vyčistil…)
- Je to hospodárné (z hlediska financování již nevynakládáme peníze na výzkum). I pokud je data nutné koupit, je to zpravidla mnohonásobně levnější než obdobný vlastní výzkum.
- Zabraňuje ztrátě cenných dat (která nevyužil původní výzkumník).
- Zdánlivě nezasahující (zpravidla již nenarušuje soukromí respondentů, mohl dříve) a nereaktivní (výzkumník neovlivňuje výsledek).

- Umožňuje pracovat s daty, která bychom nikdy sami nemohli získat (např. jsou v individuální rovině důvěrná) a nebo by to bylo velmi náročné (např. údaje o porodnosti, o všech nezaměstnaných).
- Zmenšuje problém prostorové nedostupnosti (mezinárodní srovnání)
- Data jsou mnohdy vytvořena autoritami (ČSÚ, Eurostat) = záruka + věříme v jejich kvalitu (reprezentativnost, obsahují hodně informací, velké vzorky – můžeme studovat podskupiny), cenné pro některé typy problémů.
- SA můžeme využít pro podpoření vnější validity výzkumu (srovnání našich výsledků s výsledky jiných výzkumníků, statistiky mohou být dobré pro dokreslení kontextu výzkumu)

Nevýhody a rizika sekundární analýzy dat

- Data mohou být (zpravidla jsou) původně určená k jinému záměru analýzy (chybí nám klíčové proměnné nebo informace)
- Data nemusejí být ve vhodném formátu (potřebujeme údaje o ženách, ale máme jen údaje o celé populaci, příjem je měřen v kategoriích místo v intervalu)
- Data mohou být pro ne-autora málo srozumitelná (Michael a Eaton 2012).
- Nemáme jistotu, jakým způsobem byla data skutečně získána (či nad metodikou tvorby indikátorů), v datech mohlo dojít k dodatečným úpravám
- Nevidíme v jakém kontextu byl výzkum realizován (atmosféra vyplňování dotazníků)
- Nedostupnost dat (nikdo to nesleduje nebo data nejsou publikována/veřejná), legislativní překážky, bude dlouho trvat získat povolení ke zveřejnění
- Zpravidla získáváme později než primární data (aktuálnost)
- Sekundární analýza dat – „zdrcující kritika“: příliš velký nesoulad mezi teorií a empirií, příliš široce se využívá (na nevhodné problémy), získaná data (např. publikované statistiky) jsou nespolehlivé (viz Smith 2008), z dat jsou vyvozovány neodpovídající závěry a zobecnění.

Základní otázka SA je, zda máme k dispozici individuální či agregovaná data.

Individuální data: máme informace o každém případu, které jsou zpravidla uloženy na jediném místě (v dotazníku, v adresáři, datovém souboru). V příkladu uložení v databází jsou vybrány př. proměnné jméno, pohlaví, věk a vzdělání.

Jméno	pohlaví	věk	vzdělání
Josef	muž	47 let	základní
Eva	žena	23 let	VŠ
Karel	muž	70 let	SŠ
Jarmila	žena	34 let	vyučen/a
Miroslav	muž	28 let	VŠ

- Můžeme získat datový soubor (tzv. datová matice).
- Zpravidla máme všechny údaje ke všem případům (výjimky).
- Nabízí často široké možnosti (bivariační i multivariční) statistické analýzy.
- Individuální data nejsou někdy veřejně dostupná neboť je obtížné je prezentovat, nechceme zveřejňovat osobní údaje, nutné zachovat anonymitu respondentů, prezentujeme výsledky statistických operací…

Agregovaná data:

Často jsou to data, která byla vytvořena sloučením (sumací) případů v individuálních datech do kategorií podle určitých hledisek (věk, pohlaví).

- Pak jsou data prezentována jako tzv. kontingenční tabulka.

Vzdělání / Pohlaví	Základní škola	Vyučen/a	Středoškolské	Vysokoškolské	celkem
muži	15 (Josef)	35	35 (Karel)	15 (Miroslav)	100
ženy	20	30 (Jarmila)	30	20 (Eva)	100

- např. na předchozí straně muži (Josef, Karel, Miroslav), ženy (Eva, Jarmila). Víme tedy, že ve výběru (předchozí snímek) máme celkem 3 muže a 2 ženy ze 100 mužů a 100 žen.
- Může se ale jednat také o agregaci dat o více relativně oddělených subjektech (o počtu všech studentů na vysokých školách v ČR)

Agregovaná data, která byla vytvořena složitější transformací proměnných v individuálních či v agregovaných datech. Je tedy výsledkem výpočtu. V kondenzované podobě (indexy, míry, relace).

Míra nezaměstnanosti:

počet nezaměstnaných 100

počet ekonomicky aktivních 1000

míra nezaměstnanosti 0,1 tj. v procentech (x100) = 10 %

Míry (indexy) deprivace: uvádí nakolik jsou domácnosti deprimované ve svých základních potřebách. Hodnota indexu např. 6. reprezentuje deprivaci v 6 položkách.

Odds Ratio (poměry šancí): vyjádření míry asociace mezi dvěma nominálními proměnnými. Porovnává šance na určitý stav nebo událost mezi skupinami.

	Jde na hokej	**Nejde na hokej**
Kluci	(A) 90	(B) 10
Holky	(C) 20	(D) 80

Kluci: [A:B] 9:1 (jde devět z deseti), Holky: [C:D] 0,25:1 (jde každá čtvrtá).

Odds ratio = [A:D] (0.9*0.8) / [B:C] (0.1*0.2) = 0.72 / 0.02 = 36 krát vyšší šance u kluků, že jdou na hokej.

Čím složitější a abstraktnější je indikátor, tím obtížnější je interpretace.

Typy analýzy sekundárních dat

Základní hledisko rozlišení: zda pracujeme s individuálními či agregovanými daty.

- sekundární analýza kvantitativního datového zdroje (datové matice individuálních dat). Za jednotlivé respondenty máme k dispozici všechna data a nebo pouze jejich část. Některé instituce neposkytují celá data, ale jen výtahy vzhledem k cíli analýzy + zachování anonymity.

- práce se statistikami (též terciální analýza zdrojů): práce s agregovanými daty. Ta mohou vznikat na základě individuálních dat i tzv. výkazů = instituce vyrábí tabulky (např. analýza trhu práce v regionu, vývoj počtu zařízení veřejné péče). Zpravidla musíme vystačit s tím, co již je publikováno (ale např. ČSÚ lze dotázat).

- institucionální analýza (analýza institucionálních charakteristik): analyzujeme institucionální nastavení systémů (podoba jednotlivých dávek a pravidel v jednotlivých zemích – např. zákoník práce, zákon o zaměstnanosti, zákon o státní sociální podpoře)

Specifické účely sekundární analýzy:

- Zhodnocení dosavadních zjištění o nějakém problému. Je vždy součástí tvorby textu vědecké práce (rešerše literatury), využívá se v učebních materiálech nebo jako samostatná analýza (zpravidla na objednávku).

  a) přehledová stať (state of art report) – diskutuje nějaký problém po teoretické a empirické stránce (co se o tom ví = informovat autority, navázat vlastním výzkumem, nedělat znovu to samé…).

  b) meta analýza – shromažďuje výsledky mnoha výzkumných studií na stejné téma a s těmito výsledky dále statisticky pracuje a snaží se odvodit obecnější závěr.

  \+ snaha překonat ohraničenost průkaznosti jednotlivých evidencí

  \- míchá hrušky a jablka (jiné metody, jiný kontext výzkumu)
- mezinárodní komparace: stejné statistiky nebo individuální data o více zemích.
- některé evaluace: využívá často administrativní data (např. nezaměstnanost).
- historické analýzy (vývoj sociálního státu, nezaměstnanost za první republiky, vývoj důchodového systému po roce 1990). Nelze zkoumat přímo a zpravidla ani na individuálních datech. Některé věci jsou sledovány již od 18. století.

Posouzení vhodnosti využití sekundárních dat

Klíčové hlediska pro posouzení jsou:

- Relevance z hlediska poznávacího cíle: Nakolik odpovídá využití primární nebo sekundární analýzy dat?
- Dostupné zdroje pro výzkum: čas, finanční prostředky, personální kapacity
- Dostupnost dat: fyzický přístup, cenová dostupnost, důvěryhodnost zdroje dat, stáří dat, časové omezení přístupu. Máme dostatek informací o původu dat?
- Kompletnost a vhodný formát dat: Jsou data o správné populaci a obsahují klíčové proměnné (indikátory)? Vždy je důležité snažit se získat průvodní dokumentaci výzkumu (dotazníky, popisy sběru dat). Měli bychom se seznámit s původním záměrem při tvorbě dat. Riziko „diktátu metody“ (zvolená data nám formulují výzkumné cíle).
- Posouzení etičnosti přístupu k datům (informovaný souhlas, slib mlčenlivosti, právní aspekty). Pro sekundární analýzu dat by měla platit stejná (obdobná) etická pravidla jako pro vlastní výzkum (rizikové neuvědomí si, náhodou). Data mohou někomu patřit. Za jakých okolností můžeme použít (restrikce).

  a) informovaný souhlas s využitím dat (jiný cíl může poškodit respondenta),

  b) zachování anonymity a konfidentality (ochrana citlivých údajů, např. datová matice a její uložení).

Získání dat

Zdroje sekundárních dat (co můžeme chtít využít)

- statistiky, veřejné databáze agregovaných dat sbírané pro potřeby státu a veřejné správy.
- statistické publikace (statistické ročenky)
- originální data vytvořená výzkumem jiných autorů nebo naším předchozím výzkumem (datové matice)
- výzkumné články a studie publikované na téma našeho výzkumu (nejen tabulky)
- informace o cílech a podobě systémů sociální politiky

Dostupnost: veřejné (běžně dostupné), podmíněně dostupné (se souhlasem), nedostupné (!).

Pravidla skladování dat: na bezpečném, obtížně dostupném místě.

Vést si datovou evidenci: zdroj, kde jsem data našel.

Přístup k datům

Webové stránky institucí:

Ministerstvo práce a sociálních věcí:
zaměstnanost: <http://portal.mpsv.cz/sz/stat>

Český statistický úřad:
<http://www.czso.cz/csu/redakce.nsf/i/home>

Eurostat (podle témat):
< http://ec.europa.eu/eurostat/data/browse-statistics-by-theme>
< http://ec.europa.eu/eurostat/data/database>
Dnes často možnost částečně si z dat generovat vlastní výstupy.

Česká národní banka (ARAD):
<http://www.cnb.cz/docs/ARADY/HTML/index.htm>

Mezinárodní organizace práce: (ILOSTAT)
<http://www.ilo.org/global/statistics-and-databases/lang--en/index.htm>

Statistické ročenky:

Statistická ročenka České Republiky:
< https://www.czso.cz/csu/czso/statisticka-rocenka-ceske-republiky-2015 >

Statistická ročenka z oblasti práce a sociálních věcí
<http://www.mpsv.cz/cs/3869>

Statistická ročenka trhu práce v České Republice
<http://portal.mpsv.cz/sz/stat/stro>

Ročenky školství
<http://www.msmt.cz/statistika-skolstvi/rocenky>

Zdravotnická ročenka České Republiky
<http://www.uzis.cz/katalog/rocenky/zdravotnicka-rocenka-ceske-republiky>

Bulletiny VÚPSV
<http://www.vupsv.cz/index.php?p=bulletins&site=default>

Dlouhodobá statistická šetření:

Sčítání lidu, domů a bytů (cenzus, zahrnuje všechny obyv. 1x za 10 let)
<http://www.scitani.cz/sldb2011/redakce.nsf/i/o_scitani>

Výběrová šetření pracovních sil (Labour Force Survey)
< https://www.czso.cz/csu/czso/zamestnanost-a-nezamestnanost-podle-vysledku-vsps-rocni-prumery-rok-2015 >

Příjmy a životní podmínky domácností (SILC)
< https://www.czso.cz/csu/czso/prijmy-a-zivotni-podminky-domacnosti-2014 >

International Social Survey Programme (ISSP) (každý rok jiné téma)
<http://www.issp.org/page.php?pageId=4>

European Values Study (EVS) (výzkum hodnot, např. rodina, práce…)
<http://www.europeanvaluesstudy.eu/>

Eurobarometr (mínění o veřejných otázkách)
<http://ec.europa.eu/public_opinion/index_en.htm>

Datové archivy primárních dat:

Slouží k uložení provedených šetření, která mohou být případně využita k výzkumným nebo k pedagogickým účelům.

Český sociálně vědní datový archiv (kvantitativní výzkum)
<http://archiv.soc.cas.cz/>

Systém MEDARD (kvalitativní výzkum)
< http://archiv.soc.cas.cz/data-v-archivu-medard >

UK data archive
<http://www.data-archive.ac.uk/>

Informace o sociální politice – např. MISSOC, ISSA.
<http://www.missoc.org/MISSOC/INFORMATIONBASE/COMPARATIVETABLES/MISSOCDATABASE/comparativeTableSearch.jsp >

Analýza dat

Již jsme nalezli nějaká vhodná data:

- Riziko dočasnosti přístupu k datům (získat pokud možno všechna data hned)
- Vybereme vhodná data a vyřadíme obdobná data (která máme vícekrát, z více zdrojů atp.)
- Základem je přiměřená kritika dat (zajímáme se o to jak data vznikla, kdo výzkum prováděl, studujeme dokumentaci, např. dotazníky) – velké riziko chyby na základě neznalosti dat (extrémní případ je extrapolace, kterou považujeme za data). Můžeme kontaktovat předchozí výzkumníky?
- Typ dat (agregovaná, individualizovaná) ovlivňuje možnosti analýzy.

Proč jsou individuální data pro analýzu lepší než agregovaná?

- Protože individuální data můžeme sami agregovat, ale opačný postup je možný jen ve velmi omezené míře.
- Při agregaci dat zpravidla ztrácíme část informace.
- Jsme omezení počtem proměnných v dvourozměrné tabulce (2-4).

Kritická místa výzkumu = riziko chyby v analýze nebo v interpretaci.

Dokumentace (vždy musíme sledovat při sekundární analýze)

- Popis původního projektu (k čemu byla data vytvořena, jak vypadají)
- Metadata (metainformace): Strukturovaná data o datech. Textový dokument, který obsahuje základní informace o konceptualizaci (např. mezinárodně využívané klasifikace), designu, sledovaných jednotkách (např. osoby starší 18 let), zdrojích dat, a odpovědnosti za publikaci dat.
- User guide or handbook (průvodce výzkumem): Pro konkrétní výzkumné šetření – komplexní popis šetření (cíle, metoda sběru, podmínky sběru atd.).
- Dotazník: znění jednotlivých otázek a variant odpovědi (srovnání s tím so je v datové matici.
- Codebook (Princeton University 2007): je technickým popisem konkrétního datového souboru.

  a) velikost souboru, případně jeho částí

  b) jak jsou data v matici organizována (proměnné)

  c) v jakém formátu jsou proměnné (stringové, numerické)

  d) jak jsou data kódována (muž = 1, žena = 2)

  e) které proměnné byly rekódovány a jak

  f) základní výsledky (frekvence) u jednotlivých otázek (51 procent žen….)
- Publikované výsledky: nepublikovat zbytečně totéž, konzistence s předchozími autory (ověření našeho postupu).
- Syntaxy: např. o tom jak byla data rekódována

Sekundární analýza individualizovaného datového souboru:

- Je dobré udělat si popisnou analýzu, aby se odhalily podivné (např. nemožné) hodnoty, „outliers“ a chybějící hodnoty atd. (Michael a Eaton 2012).
- Ověřit si význam hodnot: rozhodnutí o záznamu dat neděláme my, ale někdo jiný (příklad: APP = aktivizační pracovní příležitost, nepovinná statistika = nestalo se nebo nevyplnil?).
- Transformace mezi datovými jazyky. Jen někdy lze symboly převádět na čísla, opačně ještě v omezenější míře – mj. závisí na počtu variant.

  (např. název vesnice – číselník „Letiny“, „557951“) jak použít?

  čím složitější informace, tím větší je míra její redukce.
- Pozor na umělé proměnné (vznikly ne jako výsledek dotazování, ale prací předchozích výzkumníků s daty). Lépe je vycházet z neupravených dat a transformovat je sám/sama (je to i zpravidla nutnost).
- Pozor na otázky reprezentativnosti, návratnosti a případného využití vah (proměnná, která upravuje výsledky vzhledem k zastoupení určitých kategorií).

Sekundární analýza agregovaných dat

Indikátory:

- V agregovaných datech se jednotkou analýzy místo jednotlivců (případů) stávají kategorie.
- V jakých jednotkách máme údaje? (např. informace o počtu pracovníků: počet osob vs. WTE = odpracované hodiny odpovídající plnému úvazku)
- Zda máme absolutní (100, 1000) a relativní (0,1, 10 %) četnosti.
- U transformovaných indikátorů je důležité, zda je z prezentovaných dat možný zpětný proces identifikace základních, naměřených hodnot.
- Zda jsou data kompletní, které hodnoty chybí a proč.
- Problém spolehlivosti údajů v jednotlivých polích (riziko málo početné kategorie) = sledovat, zda data nejsou označena jako nespolehlivá.
- Koherence: jsou indikátory skutečně stejné? Data o některých událostech můžeme sledovat dlouhodobě (např. od 1990 do současnosti). Ale musíme si ověřit, že data byla po celou dobu sledována podle stejné metodiky. Malá změna metodiky může znamenat velkou změnu významu údaje (příklad: údaje za prosinec vs. průměrné měsíční údaje za rok).
- Problém dodatečných korekcí (tzv. průběžná data a konečná data).
- Běžně poněkud různé výsledky z různých zdrojů = jak řešit?

Interpretace

- Porozumět významu údajů. Např. pokud pocházejí z více zdrojů, časových období, zda a jak jsou standardizovaná. Některá data mohou být složitě konstruovaná z více zdrojů (např. národní účty o zaměstnanosti) =) vyhnout se chybám v interpretaci.
- Kdo data vytvořil a kdo je prezentuje (nemusí to být totožná instituce). Riziko neodborné či záměrné manipulace s daty (politici, novináři, úředníci…) a schovávání chyb v datech při jejich prezentaci (výzkumníci…).
- Záměrné zkreslení u závažných událostí jako jsou sebevraždy, násilí na ženách nebo pracovní úrazy (nenahlásí). Mnohdy záleží na definování jednotlivých kategorií (porody) = je důležité znát kontext.
- Problém s usuzováním a testováním hypotéz v terciální analýze dat. Riziko tzv. „Ecological Fallacy“ (Neuman 2007) – situace, kdy se snažíme usuzovat o menších celcích (jednotlivcích) na základě dat o větších celcích (skupinách).

  Např. v jednom městě je více bohatých lidí a více motocyklů než ve druhém, ale nemůžeme říct, že to jsou právě bohatí lidé, kdo mají motocykly (třeba tam chudí lidé mají moto-klub nebo tam je muzeum motocyklů).
- Redukcionismus – vybere dílčí věc na individuální úrovni a vysvětli makro faktory = pomíjení dalších faktorů a širšího kontextu (Neuman 2007).

Prezentace dat

Nakolik tento výstup vyhovuje z hlediska prezentace v diplomové práci?

Pohlaví * moje práce je zajímavá Crosstabulation

Count

		moje práce je zajímavá					
		rozhodně souhlasím	spíše souhlasím	ani souhlas / ani nesouhlas	spíše nesouhlasím	rozhodně nesouhlasím	Total
Pohlaví	muž	320	411	252	57	43	1083
	žena	89	144	105	25	29	392
Total		409	555	357	82	72	1475

Pohlaví * jsem celkově spokojený/á v práci Crosstabulation

Count

		jsem celkově spokojený/á v práci					
		rozhodně souhlasím	spíše souhlasím	ani souhlas / ani nesouhlas	spíše nesouhlasím	rozhodně nesouhlasím	Total
Pohlaví	muž	186	498	344	44	14	1086
	žena	107	161	107	16	6	397
Total		293	659	451	60	20	1483

Prezentace dat

- prezentace dat by měla co nejvíce odpovídat výzkumnému cíli
- musíme data upravit vzhledem k cílům našeho výzkumu (např. nepotřebujeme všechna data, ale jen jejich část, jedna tabulka může být složena z pěti jiných tabulek).
- v některých případech vytváříme nová data z dat stávajících (např. počítáme průměry, podíly, vytváříme si vlastní indikátory)
- tři typy prezentace dat: v textu, tabulky a grafy (zpravidla není nutné prezentovat všemi způsoby současně)
- v případě potřeby využíváme absolutní i relativní četnosti
- vždy označíme graf nebo tabulku názvem, případně pořadovým číslem (pro snazší orientaci)
- Identifikace datového zdroje: a) název výzkumu, b) autor dat, c) poskytovatel dat, d) umístění dat (kde jsem je získal)
- vždy jednoznačně uvádíme zdroje dat (odpovědnost, případně jaký výzkum) pod grafem či tabulkou a také v seznamu literatury (pozor na zkratky institucí jako MPSV či ČSÚ)
- je možné dopočítat si řadu údajů (i např. statistiky, nástroje na webu….)

Tabulka č. 1: Hodnocení zaměstnání respondenty v sektorech IT a strojírenství podle pohlaví (absolutně, řádková procenta)

hodnocení zaměstnání	odpověď / pohlaví	rozhodně souhlasím	spíše souhlasím	ani souhlas / ani nesouhlas	spíše nesouhlasím	rozhodně nesouhlasím	celkem
moje práce je zajímavá	muž	320	411	252	57	43	1083
		29,5 %	38 %	23,3 %	5,3 %	4 %	100 %
	žena	89	144	105	25	29	392
		22,7 %	36,7 %	26,8 %	6,4 %	7,4 %	100 %
	celkem	409	555	357	82	72	1475
		27,7 %	37,6 %	24,2 %	5,6 %	4,9 %	100 %
jsem celkově spokojený/á v práci	muž	186	498	344	44	14	1086
		17,1 %	45,9 %	31,7 %	4,1 %	1,3 %	100 %
	žena	107	161	107	16	6	397
		27 %	40,6 %	27 %	4 %	1,5 %	100 %
	celkem	293	659	451	60	20	1483
		19,8 %	44,4 %	30,4 %	4 %	1,3 %	100 %

Zdroj: výzkum „Pracovní prostředí a organizace práce" (2009).

Literatura a zdroje:

Babbie, Earl R. (2010). *The practice of social research*. (12th ed.) Belmont, CA : Wadsworth.

Blaikie, Norman W. (2003). *Analyzing quantitative data: from description to explanation*. (1st pub.) London: SAGE Publications.

Boslaught, Sarah (2007). *Secondary data sources for Public Health: A Practical Guide*, New York: Cambridge University Press.

Bryman, Alan. (2008). *Social research methods*. (3rd ed.) Oxford : Oxford University Press.

Buriánek, Jiří (1989). „K pojetí sekundární analýzy“. *Sociologický časopis*, č. 1, 1988.

Disman, Miroslav (2005). *Jak se vyrábí sociologická znalost*. Praha: Karolinum. (str. 166-176) [studium dokumentů]

Neuman, W. Lawrence (2007) *Basics of social research*. USA: Pearson Education (výběrově str. 69-107)