

II.7	Výběr části dat a třídění	468
II.8	Základní popisné statistiky	472
II.9	Složitější statistické operace	473
Dodatek III		
Přehled statistického softwaru a používání MS Excel pro analýzu dat		475
III.1	Obecné statistické balíky	475
III.2	Speciální statistický software	479
III.2.1	Speciální software pro strukturální modely	479
III.2.2	Speciální software pro víceúrovňové modely	480
III.2.3	Speciální software pro analýzu latentních tříd	481
III.3	Statistika v Excelu	482
III.3.1	Statistické funkce Excelu	482
III.3.2	Modul analýza dat	492
Dodatek IV		
Kde hledat data pro analýzu?		501
IV.1	Data z velkých mezinárodních výzkumů	501
IV.2	Datové archivy	503
IV.3	Statistická data	504
Rejstřík		505
Obsah příloženého CD		508

Úvod

Žijeme ve světě, který je prodchnut statistickými údaji. Statistická data nás doprovázejí každodenně při čtení novin, poslechu rozhlasu, sledování televizních pořadů. Citování statistických údajů velmi často sloužilo a dodnes slouží v každodenním životě jako důkaz, který má potvrdit správnost argumentace – bohužel v množství nejruznějších statistických údajů se často nacházejí i takové, které si navzájem protirečí. To může vést až k jistým pochybnostem o jejich pravdivosti, a potažmo také k pochybnostem o samotné statistice jako vědě (to je o statistické vědě), jak to naznačují dehonestující výroky typu „statistikou lze dokázat cokoli“ nebo „nevěřím žádné statistice, kterou jsem sám nezfalšoval“. I proto už v roce 1954 napsal americký žurnalista Darell Duff populární útlou knižičku s názvem *How to Lie with Statistics* (v českém překladu vyšla v roce 2013 pod názvem „Jak lhát se statistikou“), s cílem ukázat, jak se nedopouštět různých druhů chyb (a tedy statisticky nelhat) při interpretaci statistických dat.

Alespoň trochu rozumět statistice a být tak statisticky gramotný je pro každého nesmírně užitečné. Ostatně již v roce 1950 americký statistik Sam Wilks po zvolení předsedou Americké statistické asociace ve své prezidentské adrese geniálně předpověděl: „Statistické myšlení bude jednou pro efektivní občanství stejně nezbytné jako schopnost číst a psát“,¹ s čímž my, autoři této učebnice, plně souhlasíme; navíc jsme přesvědčeni, že tato doba již nastala.

Pro studenty sociálních věd je základní znalost statistických operací důležitá obzvláště: nejen proto, že jistě chtějí být efektivními občany, ale především proto, že jistě chtějí být i efektivními badateli. A jak již měli možnost v průběhu svého studia zjistit, značná část sociálněvědních závěrů a generalizujících výroků je založena právě na statistických analýzách. Studenti proto musejí být připraveni na to, že je nutné se statistiku naučit, neboť statistické operace budou organickou součástí jejich výzkumné práce. Proto musejí vědět, že statistické operace jsou založeny na určitých předpokladech, které, pokud nejsou naplněny, vedou k produkci – eufemisticky řečeno – statistických artefaktů, to je – řečeno lapidárně – k produkci mylných výsledků.

¹ Tento výrok je často mylně připisován známému anglickému spisovateli H. G. Wellsovi z jeho knihy *Mankind in the Making* z r. 1903. Wells se sice v podobném duchu vyjádřil, ale byl to Wilks, jenž Wellsovu myšlenku tak skvěle parafrázoval.

Ale i ti studenti, kteří se chtějí pohybovat především v prostředí tzv. kvalitativní metodologie, která je založena na „práci bez čísel“, by měli považovat zvládnutí základních statistických dovedností za užitečné – přinejmenším proto, aby rozuměli, jak statistické údaje vznikají a jaká čertova kopýtka se ve statistických analýzách mohou vyskytovat.

V této učebnici se budeme zabývat problematikou analýzy statistických dat prostřednictvím softwaru IBM SPSS. Považujeme za důležité hned v úvodu zdůraznit, že čtenářům nepředkládáme učebnici statistiky (proto také popisujeme základní statistické pojmy, aniž věnujeme větší pozornost tomu, jak jsou matematicky definovány), ale soubor návodů, jak statisticky analyzovat datové soubory obsahující hromadné kvantitativní údaje. Učebnice je primárně určena pro studenty společenských věd totíž oborů. Jako dlouholetí učitelé kurzů „analýza dat“ pro studenty sociálních věd totiž máme opakovanou zkušenost, že naučit naše studenty statistické analýze vyžaduje poněkud jiný přístup než prostřednictvím standardní výuky statistiky. Proto je naše učebnice napsána tak, že od čtenáře nevyžaduje více než jen základní znalosti z aritmetiky a elementární algebry. Výklad každé problematiky je řešen podle následujícího vzorce: čtenáři předestřeme analytický problém (například jaká je souvislost mezi mírou religiozity respondentů a jejich postojem k možnosti zavedení eutanazie), poté popíšeme, jakým způsobem se zadá příslušný výpočet v programu SPSS (s názornými návody) a poté ukážeme, jak je možné výsledek, který SPSS vyprodukuje, vyložit a interpretovat. Jelikož všechny analytické úlohy jsou řešeny prostřednictvím výpočtů na počítači, nemusí se nic počítat ručně. Aby ovšem čtenáři pracovali „statisticky poučeně“, výkladům některých principů statistiky se samozřejmě nevyhneme. Snažili jsme se ovšem, aby tento výklad byl maximálně srozumitelný, takže jsme museli v mnoha případech výrazně zjednodušovat (a někdy jsme se přitom dostali, jak nás ve svých posudcích upozorňovali naši recenzenti, až na tu nejspodnější možnou hranici zjednodušení).

Společenské vědy studují, jak známo, sociální jevy (fenomény), to je lidské kolektivní jednání, které je výsledkem vztahů a interakcí mezi lidmi a které se odehrává v prostředí lidské kultury a jejich organizací a institucí. Při jejich poznávání se, jako všechny vědy, řídí třemi cíli: studované jevy se 1) nejdříve musejí **popsat**, poté 2) se musejí prostřednictvím nalezení pravděpodobnostních nebo příčinných (kauzálních) vztahů **vysvětlit** a nakonec 3) je třeba se pokoušet o **predikci** (předpověď) budoucího způsobu (popřípadě variantních způsobů) jejich chování nebo existence. Jelikož sociální vědy potřebují k těmto cílům data, povolávají k jejich naplnění ve svém kvantitativním paradigmatu statistickou vědu. Ta umí prostřednictvím svých postupů, to je prostřednictvím postupů **deskriptivní statistiky**, především data **sumarizovat**, tedy **popsat**. Řekneme-li například na základě sociologického výzkumu, který byl proveden na výběrovém souboru 1 821 osob, že v roce 2008 bylo v ČR 53 % respondentů spokojeno se svým životem, zatímco nespokojeno bylo pouhých 7 %, pak jsme statisticky shrnuli 1 821 individuálních odpovědí na otázku, jak je respondent(ka) celkově spojen(a) se svým životem. Podobně sumarizující výpovědi bude, když řekneme, že v pocitu spokojenosti se muži a ženy nelišili nebo že celkově

byly v roce 2008 spokojenější se životem spíše mladší věkové skupiny než skupiny starší, neboť ve věku 18–44 let bylo spokojeno 61 % respondentů, zatímco ve věkové skupině 45 let a starších jich bylo spokojeno jen 45 %. Postupům této popisné statistiky jsou věnovány především kapitoly 3, 4 a 7.

Hledání vztahů mezi jevy s cílem nalézt jejich pravděpodobnostní nebo kauzální **vysvětlení** jsou věnovány kapitoly 8, 9, 10, 11 a 12. V kapitolách 11 a 12, jež podávají výklad jednoduché a mnohonásobné lineární regrese, se čtenář navíc seznámí s postupy, které umožňují **predikovat** budoucí chování analyzovaných jevů.

Statistická věda má ovšem pro analýzu sociálněvědních problémů ještě jeden – a to podstatný – moment. Ukazuje, za jakých okolností je možné z údajů, které sociální vědy získají z výběrových souborů (a je pro sociální vědy charakteristické, že ve svých zkoumáních pracují ne s celou populací, ale pouze s její menší či větší částí), zobecňovat prostřednictvím postupů **statistické inference** z dat těchto výběrových souborů na celou populaci (více o tom v kapitole 5).

Jak jsme již uvedli výše, naše učebnice se snaží poskytnout čtenářům pouze základní orientaci v procedurách statistické analýzy. Dobře si uvědomujeme, že tato učebnice z žádného čtenáře statistika neudělá. Věříme ale, že pomůže pochopit, co statistika je, k čemu může sloužit, jak se v ní získávají nejen přesné, ale i spolehlivé a relevantní výsledky, jak těmto výsledkům rozumět a jak je interpretovat – pozor ale, interpretace je již ze značné části za hranicemi znalostí statistiky a musí být vždy doprovázena znalostmi příslušné sociálněvědní disciplíny. Jejím hlavním cílem je tedy naučit, obecně řečeno, jak statistiku používat pro odpovědi na otázky, které si sociální vědy obecně (a sociologie specificky) kladou, a jak přitom udělat co nejméně chybných kroků a rozhodnutí nebo falešných závěrů.

Naše učebnice je výsledkem určité potřeby a z ní plynoucí poptávky, což určuje jak její obsah a výběr jednotlivých témat, tak i rozsah, který jim věnuje. Určuje ale i způsob jejich výkladu. Jsme si přitom vědomi, že se ocitáme v konkurenci se skvělými úvody do statistiky, jako jsou např. *Analýza kategorizovaných dat v sociologii* (Řehák a Řeháková, 1986) nebo *Přehled statistických metod zpracování dat* (Hendl, 2004).

Náš výklad je přizpůsoben nejen požadavkům a logice výuky statistiky v bakalářském programu sociologie, ale i logice programu IBM SPSS, který je při ní používán. V české sociologii je právě IBM SPSS momentálně nejužívanější z programů určených pro zpracování hromadných dat. I když po pravdě řečeno, z možností, které tento program pro statistickou analýzu hromadných dat nabízí, vybírá naše publikace jen malý díl. Z didaktického hlediska upozornujeme čtenáře na fakt, že pouhá četba našeho textu sice poskytne základní orientaci ve statistice, ale nenaučí jejímu praktickému použití, které je pro svou složitost vázáno na počítačové zpracování hromadných dat. K získání skutečné kompetence při práci s hromadnými daty je třeba při čtení učebnice zároveň pracovat se softwarem a uváděné příklady si krok za krokem skutečně samostatně procvičovat. Z tohoto důvodu jsou k učebnici přiloženy na CD všechny datové soubory, s nimiž se v ní operuje. Ale nejen to, doporučujeme klást