Několik poznámek k jedné obsesi českých sociálních věd – statistické významnosti*

PETR SOUKUP**
Sociologický ústav AV ČR, v.v.i, Praha
LADISLAV RABUŠIC**
Fakulta sociálních studií Masarykovy univerzity, Brno

Some Notes on the Obsession of the Czech Social Sciences with Statistical Significance

Abstract: The article explains the various errors that occur in the use of the concept of statistical significance. It points to the problem of census, nonprobability sampling, sampling of small populations and small samples. Another topic is the use of statistical methods on aggregated data files, especially from international research, and on weighted data. The authors point out that in many cases the use of statistical significance is not appropriate, and they warn against the incorrect use of traditional statistical methods. The article also presents methods that can be used to avoid the problems to which the authors have drawn attention.

Keywords: statistical significance, nonprobality samples, weighting, small proportion samples, merged data.
Sociologický časopis/Czech Sociological Review, 2007, Vol. 43, No. 2: 379–395

Aforismus o statistice: There are three kinds of lies: lies, damned lies, and statistics (Disraeli).

K sepsání tohoto článku nás přivedlo poznání, že značná část českých sociálních vědců, nemluvě o značné proporci studentů, je posedlá statistickou významností. Testy statistické signifikance v jejich povědomí (neboť tak „pochopili" smysl testování v kurzech statistiky) slouží jako všemocné zaklínadlo. Jsou přesvědčeni, že bez testů statistických hypotéz není možné získat vědecky relevantní poznatky. Domnívají se, že tyto testy musí aplikovat na všechny výsledky bez ohledu na to, zdali jejich data pocházejí z pravděpodobnostního (náhodného) výběru, vyčer-

* Práce na tomto článku byla umožněna díky grantu MPSV, číslo 1J005/04DP2 „Nerovné šance na vzdělání" (první autor) a díky grantu podpory výzkumu MPSV „Moderní společnost a její proměny" (číslo projektu 1J017/04-DP2, název projektu Vzdělávání dospělých v různých fázích životního cyklu).
** Veškerou korespondenci posílejte na adresu: PhDr. Ing. Petr Soukup, oddělení Sociologie vzdělání a stratifikace, Sociologický ústav AV ČR, v.v.i., Jilská 1, 110 00 Praha 1, e-mail: soukup@fsv.cuni.cz.

pávajícího zjišťování (z cenzu) nebo výběru nenáhodného (kvótního, záměrného, samovýběru). Jsou přesvědčeni, že testy významnosti jim řeknou, co je v datech důležitého, prostřednictvím nalezené statistické signifikance se snaží prokazovat těsnost vztahu dvou proměnných. Nic z toho ovšem statistická významnost neumí.

Cílem tohoto článku není hanět statistiku (to by asi bylo vzhledem ke vzdělání a zaměření obou autorů velmi zvláštní), ale upozornit na meze jejího používání a na její případné zneužívání zejména v (české) sociologii. V článku se věnujeme problematice používání statistických testů a podmínkám, při kterých je možné testy užívat. Naznačujeme i možnosti, jak si poradit, pokud statistické testy není možné použít, případně jak postupovat, pokud není statistické testování vůbec na místě. V neposlední řadě se krátce věnujeme i problematice práce s váženými daty, která jsou tak oblíbená, ale při práci s nimi se často chybuje. Při ukázkách se budeme snažit používat obecné postupy, někdy však uvedeme i konkrétní návod postupu v ČR zřejmě nejrozšířenějším paketu SPSS. Hned na počátku ovšem upozorňujeme čtenáře, že nelze na ploše jednoho článku popsat vše a že tento článek chápeme jako první vlaštovku, na kterou naváží další detailnější pojednání o jednotlivých problémech užívání statistické indukce.[1]

Začněme nejdřív s připomenutím samozřejmostí, na něž se ovšem v praxi často zapomíná. Jelikož teorie statistické indukce (zobecňování výsledků z výběrového souboru na základní soubor) byla primárně vyvinuta pro případy *velkých náhodných výběrů z velkých základních souborů*, není možné její běžné postupy v jiných situacích aplikovat. Použití běžných postupů statistické indukce je podmíněno čtyřmi požadavky, které si postupně uvedeme, přičemž jsme si vědomi faktu, že v praxi je nesplnění těchto požadavků zpravidla propojeno.

1. Co znamená požadavek *velkého náhodného výběru*? Jde v podstatě o tři požadavky najednou. Jednak (1) aby měl výběr dostatečný počet jednotek, jednak (2) aby bylo provedeno vybírání ze základního souboru náhodně. Posledním, ale rozhodně ne nedůležitým, požadavkem je, (3) aby šlo o výběr. Předběžně lze poznamenat, že pouze výběry, kde je vybíráno alespoň v řádu desítek jednotek cca od 30–50, lze označit za „velké".[2] Náhodné výběry jsou pak jen ty, kde o vybrání či nevybrání jednotky rozhoduje náhoda, samozřejmě ve statistickém slova smyslu, kdy náhodou rozumíme souhrn drobných, ne zcela zjistitelných či zcela nezjistitelných příčin, které způsobují, že dopředu neumíme jistě stanovit výsledek náhodného pokusu. Provedení náhodného výběru se ovšem ve vědě řídí striktními pravidly. Důležité samozřejmě je, aby

[1] Partii statistiky, která pojednává o statistické indukci, se zpravidla říká matematická statistika, v česky i anglicky psané literatuře se setkáme též s pojmy statistická inference či induktivní statistika. Neřešíme v tomto článku terminologické rozdíly mezi těmito pojmy a bereme je víceméně za synonyma.

[2] Naši studenti se nás často ptají, jak velké výběrové soubory by měli mít pro své diplomové práce. Zde je každá rada drahá, nicméně určitý návod, s nímž se ztotožňujeme, podává Blaikie [2003: 166]: 300 je adekvátní, 500 je lepší a 1000 by bylo ještě lepší.

bylo vůbec vybíráno. Pokud máme data, která pocházejí z úplného zjišťování, nemůžeme o statistické indukci vůbec hovořit.

2. A co je to *velký základní soubor*? Protože většina reálně prováděných výběrů je prováděna jako výběr bez vracení (podobně jako například ve Sportce nelze v jednom tahu dvakrát vytáhnout totéž číslo), ale statistika používaná pro statistickou indukci vychází ze vzorců výběrů s vracením, je dobré zajistit, aby základní soubor byl alespoň 100krát větší než zamýšlený výběrový soubor.[3] Není-li toto splněno, lze samozřejmě statisticky testovat, ale s jinými než běžně dostupnými vzorci. V takovýchto situacích již zpravidla nelze užívat tolik oblíbené statistické balíky, popřípadě je nutno používat jejich speciální moduly, které nejsou běžně známé a uživatel je často nemá ani zakoupeny.

Poté co jsme uvedli základní předpoklady pro používání běžných testů statistické indukce (řadíme sem zejména t-testy (jednovýběrový, dvouvýběrový, párový), analýzu rozptylu, chí-kvadrát test pro podíl (četnosti); obdobně platí naše závěry i pro užívání intervalů spolehlivosti (pro střední hodnotu, pro rozdíl dvou středních hodnot, pro podíl, resp. relativní četnost, apod.)), se budeme věnovat jednotlivým případům, kdy jsou požadavky pro její použití nesplněny, a budeme se snažit ukázat, jak v takových případech postupovat. Pro doplnění také ukážeme některé nesprávně postupy, aby bylo možno se jich vyvarovat. Ještě před detailním rozborem různých problémů statistické významnosti si zaslouží tento pojem uvedení definice a nutné je definovat i pojem reprezentativity, který je s ním nedílně spojen. V literatuře lze nalézt mnohé velmi složité definice statistické významnosti, pro účely tohoto článku volíme definici od Blahuše [Blahuš 2000: 55], protože ten zřejmě jako jediný český autor podrobnější definici nabízí.

„Tvrzení, že výsledky jsou ‚statisticky významné' na hladině a = 0,05 má přesně následující význam.

a) U náhodného reprezentativního výběru znamená, že riziko zobecnění z náhodného reprezentativního výběru na celý základní soubor je nejvýše 0,05 (tj. 5 %). Tedy např. riziko, že v základním souboru studentů není procento spokojenosti vyšší než 50 %.

Jde o riziko tzv. chyby I. druhu, že nesprávně zamítneme statistickou nulovou hypotézu H0. Tj. zde hypotézu, že rozdíl mezi skutečným procentem spokojených v základním souboru a zadaným procentem 50 % je nulový. Jinak též, že chybně zamítneme hypotézu, že rozdíl mezi hodnotou u výběru (60 %) a pesimisticky předpokládanou možnou hodnotou v základním souboru (50 %) je jen náhodný. Tedy chybně učiníme závěr, že z výběru lze provést zobecnění (zde zobecnění, že v souboru studentů je počet spokojených větší než 50 %)." Bla-

[3] Je to samozřejmě pouze orientační návod. Například pro dospělou populaci ČR ve věku 15 let a více (asi 7 miliónů osob) nemusíme mít výběrový soubor o velikosti 70 000 jednotek). A naopak, pro výzkum studentů nějaké fakulty, která má 3 000 studentů, by výběrový soubor o 30 jednotkách byl bezesporu velmi malý. Velikost výběru ovlivňují i další charakteristiky výzkumu, především hloubka analýzy, kterou chceme provádět.

huš [2000: 55] k tomu dodává, že statistická významnost tedy znamená pouze, že výsledek je „‚statisticky zobecnitelný' z reprezentativního-randomizovaného výběru na základní soubor, a to se zvoleným rizikem"[4]. Nyní se pokusme o definici reprezentativity. V běžných učebnicích statistiky pouze najdeme konstatování, že je důležité, aby výběr byl reprezentativní pro možnost induktivního usuzování z výběru na základní soubor [např. Knoke, Bohrnstedt, Mee 2002: 15]. Zřejmě jedinou cestou, jak *dosáhnout reprezentativity*, je provést *náhodný výběr ze všech jednotek základního souboru* za situace, kdy *všechny jednotky mají stejnou pravděpodobnost vybrání* [Knoke, Bohrnstedt, Mee 2002: 15]. Takto lze reprezentativitu zajistit více technikami, podrobněji se tématu věnujeme v části B.

Dále budou v článku detailněji popsány tyto situace:

A) úplná zjišťování,
B) nenáhodné výběry,
C) malé výběry,
D) výběry z malých populací,
E) velké výběry, agregace dat, mezinárodní datové soubory a
F) používání vah u náhodných výběrů.

A) Úplná (vyčerpávající) zjišťování (cenzy) a používání statistické indukce

Pokud neprovádíme výběr, ale máme informaci o všech jednotkách základního souboru, pak samozřejmě nepotřebujeme usuzovat na situaci v základním souboru a nepotřebujeme statistickou indukci. Předpokládejme, že jsme provedli šetření, v němž byl zjišťován průměrný čistý měsíční příjem ekonomicky aktivních obyvatel ČR. Zjistili bychom, že příjem mužů je vyšší než příjem žen. V takové situaci jde o skutečný rozdíl, není třeba žádného testu, abychom tuto skutečnost prokázali. V praxi je provádění úplných šetření málo časté, ale u malých základních souborů k němu může odcházet. Při používání statistické indukce bychom dospěli mnohdy k závěrům o *statisticky nevýznamných* rozdílech, ač o rozdíly ve skutečnosti jde.

S fenoménem úplných zjišťování jsou spojeny ještě dva dílčí problémy. První nastává, pokud provedeme úplné zjišťování, ale z nejrůznějších důvodů nezískáme informaci o všech jednotkách ve výběru. U šetření založených na dotazování hovoříme o míře návratnosti (*response rate*). Někteří autoři jsou toho názoru, že podmínkou aplikace statistické indukce je i vysoká návratnost – Blaikie [2003: 167] hovoří o 85 %, byť jedním dechem dodává, že v současnosti je naplnění tohoto požadavku velmi obtížné (v českých pravděpodobnostních výzkumech se pohybuje mezi 40–60 %). V této situaci rozhodně nelze postupovat tak, že budeme náš soubor považovat za výběr (navíc náhodný) ze základního souboru a budeme

[4] Blahuš podává definici i pro případ náhodných experimentů, ty jsou ovšem v sociologii řídké, proto tuto část definice neuvádíme.

používat statistickou indukci. Toto pravidlo platí i v případě, že se nám podaří získat informaci i například jen o polovině jednotek. A co v tomto případě lze dělat? Nutné je za pomoci analýzy zjistit, z jakých důvodů a případně kdo neodpověděl. Poté je možné za pomoci postupů obdobných vážení (viz bod F) se pokusit napravit možná zkreslení. Obdobný problém nastává v případě, kdy máme malý základní soubor (řádově desítky až stovky jednotek), ale z určitých důvodů se rozhodneme nezjišťovat informace o všech jednotkách. Tento případ popisujeme v bodě D) a vysvětlujeme, jak postupovat při využívání statistické indukce.

Druhým problémem nebo spíše zapeklitostí je možnost zkoumání některých vztahů pomocí statistických testů u úplných zjišťování. V případě, že chceme zjišťovat, zda spolu souvisí dvě nominální proměnné, můžeme samozřejmě sestrojit kontingenční tabulku a spočítat koeficient kontingence. Z tabulky a hodnot koeficientu pak víme, zda proměnné spolu souvisí, a případně, zda je souvislost výrazná (těsná), či nikoliv. Obdobně u dvou pořadových proměnných můžeme spočítat Kendallovo Tau (nebo stále populárnější koeficient gamma), u kvantitativních proměnných pak Pearsonův korelační koeficient (pro předpokládané lineární vztahy). Namístě již ale nejsou testy, zda příslušný koeficient je nulový v základní populaci, či nikoliv. Například pokud vyjde z úplného zjišťování korelační koeficient mezi příjmem a počtem let vzdělání 0,48, jde o středně silnou souvislost a žádný test o nulovosti korelačního koeficientu nepřichází v úvahu. Korelační koeficient o velikosti 0,48 zjištěný ze základního souboru prostě nulový není.

Je tedy důležité ještě jednou připomenout, že v případě vyčerpávajících šetření není namístě užívat statistickou indukci vůbec.

B) Nenáhodné výběry

Jak již bylo uvedeno na počátku, byla statistická indukce a její jednotlivé postupy vyvinuty pro případ náhodných výběrů (nejdříve zejména pro nejjednodušší případ prostého náhodného výběru). Pro tento případ platí veškeré vzorce pro intervalové odhady, testy statických hypotéz. Co ale dělat, pokud nemáme náhodný výběr? Nejdřív si vůbec stručně připomeňme, jaké výběry můžeme v praxi provádět. Zpravidla se užívá následující dělení výběrů:[5]

náhodné	úsudkové/nenáhodné
- prostý	- kvótní
- vícestupňový	- záměrné atd.
- oblastní apod.	

[5] Podrobný popis jednotlivých variant výběrů přesahuje možnosti tohoto textu, zájemce odkazujeme zejména na texty Čermáka a Vrabce [Čermák, Vrabec 1998a, 1998b, 1999] a na Thompsona [Thompson 2002] nebo klasický text Kishe [Kish 1965].

Z uvedeného schématu vyplývá, že v praxi tolik oblíbený *kvótní výběr není výběrem náhodným*, a tudíž statistická indukce nemá při jeho používání místo. Samozřejmě expert, který provádí kvótní výběr, může na základě svých znalostí a zejména zkušenosti být schopen stanovit chybu takového výběru a používat tak „obdoby" induktivních postupů. V praxi se však často bezhlavě (a obáváme se, že i bez vědomostí) na výsledky z kvótního výběru aplikují postupy statistické indukce. Což je samozřejmě špatně. Pochopitelně si nemyslíme, že je špatné používat kvótní výběry (spíše naopak), špatné je však u dat z takovýchto výběrů užívat induktivní statistiku bez rozmyslu.

V souvislosti s typy výběrů je nutné připomenout, že náhodným výběrem není ani *samovýběr*, tj. situace, kdy výběr jednotky závisí na jednotce samé a nikoliv na náhodě (příkladem nechť je anketa). Z tohoto důvodu u samovýběrů nemá induktivní statistika místo; dokonce žádné zobecňování na základní soubor nemá u samovýběrů místo!

Poslední poznámku k náhodnému výběru činíme z praktických důvodů. Je velice jednoduché teoreticky říkat, že nejlepší je náhodný výběr, protože pak lze užívat induktivní statistiku se všemi jejími kouzly, a odmítat všechny nenáhodné výběry. V praxi je však velice problematické náhodný výběr provést. Jak získat oporu výběru (tj. seznam všech jeho jednotek), chceme-li například provádět výběr celé dospělé populace v ČR? Shromažďování osobních údajů podléhá zákonu č. 101/2000 Sb., o ochraně osobních údajů, a je nadto nesmírně obtížné. Provést výzkum, kde bychom chtěli takovou oporu získat, by mohlo trvat také několik let. V praxi se nadto při realizaci náhodných výběrů (zpravidla vícestupňových) setkáváme s velkým podílem odmítnutých rozhovorů. Můžeme pak výběr, kde nám polovina respondentů odmítne, považovat za náhodný? Domníváme se, že nikoliv.

Jednou z možností, jak bez opory výběru provést náhodný výběr, je technika *náhodné procházky* (*random walk*),[6] která je v praxi občas užívána a nelze proti ní mít výrazné námitky. Jen ve stručnosti popišme, jak může probíhat. Tazatel vyráží z náhodně vybraného místa, například křižovatky ve městě. Má například určenu cestu, první ulici vpravo, pak třetí doleva a zde do třetího domu, 1. patra. Zde náhodně vybere byt a v tomto bytě dotazovanou osobu například metodou prvních narozenin. Obdobně pokud je malý základní soubor a těžko dosažitelný, lze použít *techniku sněhové koule* (*snow ball technique*). Ta spočívá v tom, že vybereme prvního respondenta a on nám sám doporučí dalšího. Jisté nebezpečí zde spočívá v tom, že některé charakteristiky jednotek výběru budou systematicky zkresleny, neboť lidé doporučují ty, které sami znají a kteří jsou (tak trochu) jako oni sami.

[6] Nezaměňujme toto označení se stejně označeným stochastickým procesem při modelování časových řad, i když podobnost lze jistě nalézt.

C) Malé výběry

Vzorce statistické indukce používané běžně ve statistice (a nadto vyučované běžně ve statistických kurzech) vycházejí z předpokladu, že výběr (resp. i jednotlivé podskupiny,[7] za které děláme závěry) má minimálně 30–50 jednotek. Samozřejmě může dojít k situaci, kdy takový výběr nemáme a v praxi zejména u specifických podskupin k takové situaci nezřídka dochází. Důvodem této skutečnosti mohou být často peníze, protože zadavatel výzkumu rád získává co nejvíce informací za co nejméně peněz (na tom není nic špatného), a nutí tím výzkumníky snižovat velikost výběru. Nicméně použít v tomto případě klasické postupy statistické indukce není správné.

Jaké alternativy se nám nabízejí, chceme-li přeci jen čerpat z hluboké studnice poznatků induktivní statistiky?

1) Užití *speciálních testových statistik* (nebo jednodušeji testů[8]) vyvinutých pro malé výběry.
2) Užití *neparametrických metod* s „přesnými" hodnotami testových kritérií.[9]

Ad 1) Přístupy založené na speciálních statistikách, odvozených pro malé výběry jsou spíše „hrátkami" statistiků a každý výpočet založený na těchto statistikách by bylo nutné provádět s kalkulačkou v ruce (popřípadě si naprogramovat vlastní proceduru), protože tyto postupy nejsou zahrnuty ve standardních statistických paketech. Z tohoto důvodu nelze tuto cestu pro zpracování větších úloh a počítání více analýz doporučit. Zájemce o tyto postupy lze odkázat na texty [Kahounová 2000; Řehák, Řeháková 1986: 62, 120].

Ad 2) Použití neparametrických metod s „přesnými" hodnotami testových kritérií je zřejmě v případě malých výběrů praktičtější variantou. Nutno poznamenat, že pro větší výběry má testové kritérium neparametrických testů (resp. funkce odvozená od tohoto testového kritéria) nejčastěji přibližně rozdělení normální nebo jiné běžně užívané rozdělení. Testování u větších výběrů probíhá porovnáním testového kritéria s kvantily běžně užívaných rozdělení. Pro malé výběry tato konvergence k běžně známým rozdělením neplatí a existují speciální tabulky [Anděl 2003, 2005; Blatná 1996; Likeš, Laga 1978], v nichž lze nalézt hodnotu, s níž porovnávám vypočtené testové kritérium. Také statistické pakety zohledňují tento přístup a za pomoci simulačních metod nebo jiných přístu-

[7] Z těchto důvodů se zpravidla design výzkumu, zejména velikost výběru, odvíjí od požadavku na dostatečně velký počet respondentů ve skupinách, za něž mají být samostatně zjišťovány výsledky, popřípadě má dojít ke srovnání s výsledky jiných skupin. Bývá zvykem, že pro vyšší „kvalitu" výsledků se minimální vzorek pro jednotlivé skupiny stanovuje na cca 80–100 respondentů.

[8] Každý běžný statistický test je založen na testové statistice, její hodnota se vypočítá z výběrových dat a srovná se s kvantily příslušného statistického rozdělení a učiní se závěr.

[9] Důležitá nejsou až tak přesná testová kritéria, ale důležité je, že se u malých výběrů neužívá aproximace testových kritérií za pomoci nejběžnějších rozdělení (jako je zejména normované normální), ale využívá se simulačních postupů (resp. speciálních tabulek) k nalezení přesnějších hodnot, s nimiž má být porovnáno testové kritérium.

pů umožňují provádět přesné testování. Zřejmě nejběžnější statistický software v ČR SPSS má samostatný modul nazvaný *Exact Tests*, který slouží k testování statistických hypotéz u neparametrických testů v případě malých výběrů. Jen pro připomenutí uveďme, že neparametrické metody slouží k testování hypotéz zejména v případech, kdy proměnné nemají požadovaný charakter (zejm. nejsou kardinální normálně rozdělené). Daní, kterou platíme za přechod od požadavku na kardinální normálně rozdělené proměnné k ordinálním proměnným, je nižší síla neparametrických testů (schopnost zamítnutí testované hypotézy za situace, že tato ve skutečnosti neplatí) oproti jejich parametrickým protějškům [Hendl 2004]. Známe neparametrické obdoby t-testů (jedno-, dvouvýběrového a párového) a analýzy rozptylu, korelačních koeficientů apod. Pro ilustraci špatného a správného testování rozdílu ve středních hodnotách uveďme příklad na srovnání průměrného příjmu mužů a žen v případě malého výběru (tabulka 1).

Tabulka 1. Rozdíly platů mužů a žen v malých výběrech[10]

Nesprávně použitý t-test poskytne tyto výsledky:

Dvouvýběrový t-test s nerovností rozptylů

	muži	ženy
Stř. hodnota	8644,615	6572,222
Rozptyl	6523877	8359444
Pozorování	13	9
t stat	1,73262	
P(T<=t) (1)	0,051196	
t krit (1)	1,745884	
P(T<=t) (2)	0,102391	
t krit (2)	2,119905	

Ženy ($n_1 = 9$), muži ($n_2 = 13$), vypočteno v Excelu
Zdroj: vlastní výpočty, náhodný výběr z dat ISSP 1999, n=22.

Závěr zní: *nulovou hypotézu nelze zamítnout*, nelze říci, že průměry příjmů mužů jsou vyšší než příjmy žen. Vypočtenou významnost 0,102 je nutno dělit dvěma, abychom získali jednostranný test, neboť naše alternativní hypotéza byla směrovaná (directional). Tedy 0,102/2 = 0,051. Podívejme se na výsledky Mann-Whitney testu dle přesných kritérií. Počítáme dle vzorců [Anděl 2003: 103]:

$U_1 = n_1 * n_2 + n_1 * (n_1-1)/2+T_1$,
$U_2 = n_1 * n_2 + n_2 * (n_2-1)/2+T_2$,
kde T_1, resp. T_2 je součet pořadových čísel hodnot pro ženy, resp. muže.

[10] Data použitá pro tento příklad jsou k dispozici na vyžádání u prvního autora.

Testové kritérium se určí jako menší z hodnot U_1 a U_2 a srovnává se s tabulkovými hodnotami. V našem případě je hodnota $T_1 = 72$ a $T_2 = 181$, z toho dopočteno $U_1 = 54$ a $U_2 = 27$. V tabulkách [Blatná 1996: 203; Anděl 2003: 266] můžeme zjistit, že pro $n_1 = 9$ a $n_2 = 13$ je kritickou hodnotou pro jednostranný test na 5% hladině významnosti hodnota 33, pro 2,5% hladinu významnosti pak 28. *Nulovou hypotézu lze zamítnout i na 2,5% hladině* významnosti. *Můžeme vidět, že správná metoda (neparametrická přesná) dává zcela jiné výsledky než nepřesná metoda parametrická.*

Dodejme, že statistické programy umožňují často tyto přesné výpočty (například modul Exact v SPSS), případně za pomoci simulací (metodou Monte Carlo) lze stanovit interval spolehlivosti pro hladinu významnosti.

D) Výběry z malých populací

V některých výzkumech (zejména akademického charakteru) narážíme na skutečnost, že náš základní soubor je poměrně malý (pod tímto pojmem máme na mysli řádově stovky osob). Samozřejmě že v této situaci by bylo optimální udělat úplné zjišťování, ale to často není z finančních a časových důvodů možné. Přistoupí-li výzkumník k rozhodnutí, že z malé populace udělá výběr (často vzhledem k základní populaci dost velký) a chce zároveň používat postupy statistické indukce pro usuzování na celou populaci, je potřebné modifikovat běžně užívané postupy. Ještě než si přiblížíme tento postup podrobněji, věnujme se krátce problematice výběrů z malých populací.

V případě, že je prováděn výběr z malé populace, je ještě důležitější než v případě výběrů z velkých souborů, aby se jednalo o náhodný výběr. Jakékoliv samovýběry a jejich obdoby je nutno u malých základních souborů naprosto jednoznačně odmítnout. Problémem u výběrů z malých souborů (jejichž základní charakteristiky nejsou zpravidla známy) je skutečnost, že nelze posoudit reprezentativitu. Proto naléháme na požadavek striktně provedeného náhodného výběru, který by reprezentativitu měl zaručovat.

Nyní se opět vraťme k tomu, jak vypadá modifikace statistické indukce v případě výběrů z malých souborů. Předpokládejme, že stojíme před následujícím problémem: Byl proveden výběr o velikosti 150 (dále ve vzorcích symbol n = 150) ze základního souboru o velikosti 300 (dále ve vzorcích N = 300). Problémem, na který narazí výzkumník používající běžnou statistickou indukci obsaženou ve statistickém softwaru, bude skutečnost, že při použití dvouvýběrového testu (nebo analýzy rozptylu v případě více skupin) vychází rozdíly mezi skupinami jako nevýznamné, intervaly spolehlivosti (pro střední hodnoty a relativní četnosti) jsou poměrně široké atd. Důvodem je skutečnost, že vzorce pro klasickou statistickou indukci vycházejí z předpokladu, že se provádí výběr s vracením,[11] i když

[11] Postup, kdy jakoby pomyslně vybíráme z osudí a jednotka (např. respondent) po vybrání je vrácena zpět do osudí a může být vybrána i vícekrát.

v praxi se používá téměř výhradně výběr bez vracení.[12] Vychází se z poučky, že *konečnostní násobitel* (finite population correction factor), kterým se liší[13] vzorec rozptylu u výběru bez vracení a výběru s vracením, se v případě výběru z velké populace blíží svou hodnotou 1 a lze říci, že rozptyl u výběrů s vracením a výběrů bez vracení z velkých populací je v podstatě shodný (a využívá se pro výběry bez vracení jednodušších vzorců pro výběry s vracením). Tento poznatek ale neplatí u výběrů z malých populací, kdy vybíráme podstatnou část základního souboru (i v námi uvedeném příkladě, kde $n/N = 1/2$, tedy je vybrána jedna polovina základního souboru) a rozptyly výběrů s vracením a bez vracení se liší. A jak vypadá vzorec konečnostního násobitele?

$$K = \frac{N - n}{N - 1},$$

kde N je velikost základního souboru a n velikost výběru.

Vypočteme-li hodnotu K v námi uvedeném příkladu, dosazením do vzorce získáme hodnotu přibližně rovnou 1/2. Touto hodnotou musíme korigovat rozptyl v klasických vzorcích statistické indukce, v našem příkladě rozptyl snížíme o jednu polovinu. Ukažme si na našem smyšleném výběru 150 respondentů ze 300 na příkladu užití konečnostního násobitele.

Příklad: Určete 95% intervalový odhad podílu (relativní četnosti) osob, které byly ve vězení déle než 10 let, když z výběru jste získali bodový odhad relativní četnosti hodnoty p = 0,3 (resp. po vynásobení jedním stem 30 %). Připomeňme, že máme základní soubor o velikosti 300 a z něj vybíráme 150 osob.

Výpočet

Nejdříve počítejme, jak je běžné:

Intervalový odhad rel. četnosti = $p \pm u_{0,975} * \sqrt{(p * (1 - p)/n)}$, kde $u_{0,975}$ je 97,5% kvantil normovaného normálního rozdělení (který jak známo má hodnou přibližně 1,96).

Výsledný intervalový odhad po dosazení do vzorce je mezi 0,23–0,37 (tedy mezi 23 % a 37 %).

Jelikož ale máme výběr z malé populace, musíme korigovat rozptyl konečnostním násobitelem. Rozptyl je ve vzorci vyjádřen výrazem pod odmocninou, vzorec výše uvedený můžeme korigovat násobením odmocninou z konečnostního násobitele, pro úplnost uveďme celý vzorec:

Intervalový odhad rel. četnosti = $p \pm u_{0,975} * \sqrt{(p * (1 - p)/n)} * \sqrt{K}$

[12] Postup, kdy jakoby pomyslně vybíráme z osudí a jednotka (např. respondent) po vybrání není vrácena zpět do osudí a nemůže být vybrána i vícekrát.

[13] Přesnější je vyjádření, že konečnostní násobitel ukazuje, kolikrát se liší rozptyl u výběru bez vracení oproti výběru s vracením.

Výsledný intervalový odhad je mezi 0,25 a 0,35 (25–35 %), je užší. Konkrétně je užší násobkem odmocninou z jedné poloviny (konečnostního násobitele). V našem případě činí zhruba 70 % původního intervalu.

Poznamenejme, že obdobný postup můžeme uplatnit i pro interval spolehlivosti pro průměr (například pro průměrný příjem domácnosti), popřípadě pro rozdíl mezi průměry dvou skupin (výběrů) a pro klasické testy srovnávající průměry (t-testy).

Zobecníme-li výše uvedené, máme-li výběr z malého základního souboru, spočteme dle výše uvedeného vzorce konečnostní násobitel, resp. odmocninu z něj. Poté postupujeme následovně:

1) V případě intervalového odhadu zkrátíme interval vynásobením odmocninou z konečnostního násobitele. Náš příklad byl ukázán na oboustranném intervalovém odhadu, ale samozřejmě korekci odmocninou konečnostního násobitele můžeme použít i u jednostranných intervalových odhadů.
2) V případě statistických testů vynásobíme testové kritérium převrácenou hodnotou odmocniny z konečnostního násobitele, zvýší se hodnota testového kritéria. Tu pak budeme muset porovnat s hodnotou kvantilu příslušného statistického rozdělení a učinit závěr o zamítnutí-nezamítnutí testované hypotézy.[14]

E) Velké výběry, agregace dat, mezinárodní datové soubory

V některých případech je výzkumník ve zdánlivě dobré situaci, protože má k dispozici velký výběrový soubor. V případě výběrů v řádu tisíců pak již na první pohled téměř u všech charakteristik vychází významné rozdíly (které mohou být například u průměrů na 5bodových škálách na úrovni menší než 0,1), téměř všechny závislosti měřené nejrůznějšími koeficienty jsou významné apod. I taková je odvrácená tvář statistické indukce. Výzkumník může mít radost, že nalézá rozdíly a závislosti, ale je opravdu právě toto zjištění jeho cílem? Nemá jít spíše než o statisticky významné rozdíly o rozdíly věcné navíc určité velikosti? Samozřejmě že ano, a proto cílem tohoto oddílu má být varování před svody, že výsledky jsou automaticky dobré, pokud vychází jako statisticky signifikantní.

Ukažme si, jak lze relativně uměle dosáhnout velkých výběrů včetně popsaných efektů typu „vše souvisí se vším, každý rozdíl je významný". V komerční praxi často užívané (u tzv. kontinuálních výzkumů) spojování dat může přinést kýžené výsledky. Provádíme-li měření za pomoci stejného dotazníku každý týden, měsíc apod., není nic snazšího než data z různých okamžiků spojit a začít testovat výsledky na spojených datech. Problém je, že měření z různých časových okamžiků mohou být zatížena různými chybami, které nevědomky

[14] Staromilci budou hledat v tabulkách, pokrokáři pak v tabulkových kalkulátorech anebo ještě lépe ve statistických paketech, kde dokonce mohou využít luxusu hledání hladiny významnosti, na které je ještě přípustné zamítnout testovanou (nulovou) hypotézu (to, co se značí jako Sig., P, P-level apod. v běžných výstupech).

sčítáme, takže pak zdaleka neplatí pravidlo statistické indukce o poklesu chyby s nárůstem velikosti výběru. Dalším přítomným fenoménem samozřejmě může být časový vývoj sledovaného ukazatele pouze v některé ze sledovaných skupin apod. Otázkou také je, co nám říká například výsledek o odlišnosti spokojenosti mužů a žen s jejich mobilním telefonem na datech spojených za poslední tři roky. Proto než bezhlavě testovat na spojených datech za dlouhá časová období je často lepší testovat spíše na kratších úsecích například neparametrickými metodami.

Obdobným nešvarem s možná ještě horšími výsledky je testování na datech spojených z několika „různých" výzkumů. V zásadě se nabízejí tyto možnosti:

1) data z jednoho výzkumu, která byla sebrána několika institucemi (příkladem v ČR může být Media projekt),

2) data z jednoho výzkumu sbírána v různých zemích (příkladem může být projekt EVS, WVS, ISSP apod.),

3) data spojená ad-hoc výzkumníkem z několika projektů, které obsahují tytéž otázky (často měřené v nejrůznějším čase).

Ponechme teď stranou variantu 3, která se podobá variantě uvedené v předchozím odstavci (částečně ale i variantám 1) a 2)) a zaměřme se na varianty 1) a 2), jež jsou si v mnohém podobné. Co je problémem takového spojování? Předně skutečnost, že zpravidla každý subjekt, který sbírá data, se dopouští určitých systematických chyb, jež se při spojování dat mohou jen umocnit. Navíc v případě nekvality jednoho ze spojovaných datových souborů je ihned tato nekvalita zanesena do celkových dat. Ještě větší problém souvisí s různými postupy při výběrech v různých organizacích a případném následném vážení. Zejména v mezinárodních projektech dochází často k situaci, kdy výsledný soubor má u některých zemí váhovou proměnnou, u jiných ji však nemá.

Výhodou spojených souborů je samozřejmě efekt výše popsaný, tedy vše na sobě závisí, vše se od sebe liší. Ale má například opravdu smysl testovat rozdíly mezi muži a ženami na evropské/celosvětové úrovni? Osobně se domníváme, že daleko lepší je provést dílčí analýzy na národních úrovních a pak postupy metaanalýzy a dospět ke všeobecným závěrům. Při práci se spojenými daty je také zapotřebí víc než kdy jindy kontrolovat, zda jsou data v pořádku, a to nejen na úrovni spojeného souboru, ale zejména na úrovni jednotlivých spojovaných souborů. Ze všech těchto důvodů proto varujeme před podlehnutím kouzlu spojování souborů a před svodem laciného získávání statisticky významných výsledků, které jsou však ve skutečnosti dosti *bezvýznamné*.

Aby bylo ukázáno, že statistika myslí i na případy, které zde popisujme, uveďme si, že *s daty spojenými za relativně homogenní skupiny* (ze něž lze považovat i jednotlivé země) *lze pracovat za pomoci víceúrovňových modelů* (hierarchických modelů[15]). Používání těchto modelů zatím není v české sociologii samozřejmos-

[15] Anglická terminologie užívá pojmů multilevel modelling, hierarchical modelling, ale např. v ekonometrické literatuře random-coefficient model apod. Přehled nejrůznějších pojmů pro tyto modely lze nalézt v Raudenbusch, Bryk [2002: 5–6].

tí [Soukup 2006; Hamplová 2005 a Hendl 2004], více nalezneme v zahraniční literatuře [např. Hox 1995, 2002; Norušis 2004; Raudenbusch, Bryk 2002].

F) Používání vah u náhodných výběrů

Než si ukážeme, kdy se nesprávně používá statistická indukce na vážených datech, zkusme se krátce zamyslet nad tím, kdy k vážení dochází a zda je třeba tuto proceduru užívat. Pokusme se nalézt společné rysy případů, ve kterých výzkumník zkouší vytvořit váhu a posléze ji aplikuje na svá data a pracuje s váženými daty. V praxi se nejčastěji vyskytují tyto dva případy:

1) úprava výběru takovým způsobem, aby jeho vybrané (zpravidla demografické) charakteristiky (resp. proporce z hlediska těchto charakteristik) odpovídaly hodnotám (proporcím) v základním souboru.
2) Spojení souboru ze „základního" výběru s dodatkovým výběrem (tzv. boostem).

Samozřejmě, že někdo může namítnout, že případ 2) je speciálním případem varianty 1) (a někdy i jejím důsledkem), ale pro jeho odlišnou logiku si o něm řekneme zvlášť.

Ad 1) Jak vypadá tvorba váhy u 1. varianty?[16] Vše demonstruje tabulka 2:

Tabulka 2. Ukázka vážení dat podle jedné proměnné

	ZŠ	SŠ bez maturity	SŠ s maturitou	VŠ	Součet
I. Struktura výběru	21 %	40 %	35%	4 %	100 %
II. Struktura základní soubor (ČSÚ)	23 %	41,5 %	28 %	7,5 %	100 %
III. váha pro jednotlivé skupiny (II/I)	*1,095*	*1,0375*	*0,800*	*1,875*	×
IV. (I * III) váha * počet ve výběru	23 %	41,5 %	28 %	7,5 %	100 %

Zdroj: vlastní smyšlený výběr, proporce základního souboru ČSÚ 1999.

Výzkumník stanoví váhy pro jednotlivé respondenty dle jejich vzdělání, tak aby po spuštění váhy (při práci s váženými daty) odpovídaly proporce údajům z ČSÚ. Až potud je postup zcela správný. Samozřejmě za předpokladu, že výběr byl opravdu náhodný, o čemž se lze přesvědčit zejména za pomoci chí-kvad-

[16] Omlouváme se za poněkud triviální výklad této problematiky, ale z edukativních důvodů uvádíme problematiku práce s váženými dat popisem procesu vzniku vah. Čtenáře znalé těchto postupů prosíme, nechť přeskočí tyto triviální popisy a pokračují četbu dalším odstavcem odhalujícím úskalí práce s váženými daty postupy statistické indukce. Totéž platí i pro dále uvedený popis postupu ad 2).

rát testů dobré shody, které umožňují testovat hypotézu o náhodném vychýlení struktury ve výběry [Herzmann et al. 1995: 93; Anděl 2005: 271; Zvára, Štěpán 2002: 195].

Pokud bychom statistickou indukci používali jen pro zobecnění výsledků za celek na celou populaci, byla by práce s váženými daty v pořádku. Problém ale nastává v okamžiku, když například budeme v našem případě zkoumat rozdíly mezi vzdělanostními skupinami. Budeme-li v takovém případě pracovat s vahou, počítáme s jinými (umělými) počty respondentů v jednotlivých skupinách (například u vysokoškoláků cca 1,9krát vyššími) a výsledky statistických testů najednou mohou být významné jen díky tomuto umělému navýšení. Namístě je testování bez spuštěné váhy. Docházíme tak k zesložitění práce s váženými daty, pro jednu úlohu váhu užijeme, pro jinou nikoliv. Situace je v praxi ještě komplikovanější, protože váha se zpravidla nestanoví jen za pomoci jedné charakteristiky (proměnné), ale za pomoci více charakteristik. Opomíjíme složitý praktický problém, kde získat sdružené distribuce těchto více charakteristik za celou populaci (každý, kdo se o to pokoušel, o tom ví své). I z těchto důvodů bychom rádi problematice vážení a možností statistické indukce u vážených dat věnovali samostatný navazující článek.

Ad 2) Věnujme ještě krátce pozornost situaci, kdy jsme provedli výběr a poté ještě dodatečný výběr[17] například jen jedné skupiny (například osob s VŠ vzděláním). Motivací těchto dodatečných výběrů je fakt, že určité skupiny jsou málo zastoupeny, a proto uměle navýšíme jejich zastoupení ve výběru. Jak konstruujeme váhu v tomto případě? Zjistíme strukturu z hlediska relevantní charakteristiky v základním výběru (který musí být samozřejmě náhodný a reprezentativní) a váhu stanovíme tak, aby tato struktura byla zachována i na datech po sloučení základního a dodatečného výběru. Zatímco v případě označeném 1) jsme na výběr aplikovali strukturu základního souboru, nyní aplikujeme na spojená data ze dvou výběrů struktur výběrového souboru. Samozřejmě tomuto kroku může předcházet aplikace postupu dle 1) na základní výběrový soubor, nicméně toto „dvojí" vážení spíše nedoporučujeme.

A nyní opět zauvažujme nad užitím statistické indukce u takto vážených dat. I zde platí, že chceme-li usuzovat na výsledky za celou populaci, užijeme váhu (srovnatelné výsledky bychom měli získat i v případě, že použijeme data bez váhy bez dodatečného výběru). V případě, že chceme testovat rozdíly mezi vzdělanostními skupinami, měli bychom opět pracovat s neváženými daty (i když v tomto případě bychom se nedopustili takové chyby, jako v případě 1), neboť u skupin zahrnutých v dodatečných výběrech bychom pracovali s nižšími počty respondentů a méně rozdílů by bylo statisticky významných, což je jistě méně nebezpečné).

[17] Zde narážíme na meze české terminologie výběrových šetření, logicky by se nabízel pojem dovýběr, ale to je pojem užívaný pro doplňování počtu respondentů v případě, že máme méně respondentů, než jsme zamýšleli a rozhodujeme o něm ex-post. Dodatečný výběr málo zastoupených skupin je naopak zvolen zpravidla již na počátku výzkumu (tedy ex-ante) a provádí se společně se základním výběrem. Z tohoto pohledu není možná adjektivum „dodatečný" přiléhavé a bylo by lépe užít adjektivum „doplňkový".

Opět uveďme, že v praxi se situace zpravidla komplikuje, protože není prováděn jeden dodatečný výběr, ale je jich prováděno více. Vážení je v takovém případě logicky složitější a práce s váženými/neváženými daty samozřejmě také.

Shrnutí poznatků a výzvy do budoucna

Poté co jsme nastínili možné konkrétní problémy, ke kterým může docházet při nesprávném mechanickém užívání klasických postupů statistické indukce, ještě dodejme, že problémy spojené se statistickou významností tímto nekončí. V metodologické literatuře najdeme mnoho výtek proti konceptu statistické významnosti, radikální autoři dokonce navrhují přestat tento koncept užívat. Jsme si vědomi, že není možné v jednom článku tuto diskusi plně postihnout, proto připravujeme článek, který bude doplňkem tohoto textu a jehož cílem bude ukázat na obecné meze statistické významnosti a také na koncepci věcné významnosti a jejího měření.

Závěrem jen poznamenejme, že věcná významnost a možnosti jejího měření zatím není standardním obsahem statistických, ale ani obecnějších metodologických učebnic. Nicméně to není důvodem, abychom se tématu dále nevěnovali, nebo ho dokonce zcela vytěsňovali z metodologických diskusí.

Závěr

V tomto článku jsme se zabývali problematikou statistické indukce a jejích možností. Vyšli jsme z našich pedagogických i výzkumnických zkušeností, které ukazují, že tato problematika je jednou z nejhůře pochopených pasáží statistiky. Četba zahraničních učebnic analýzy dat pro sociální vědce naznačuje [viz např. Blaikie 2003; de Vaus 2002; Field 2005], že v tom nejsme (naštěstí) tak úplně sami – a Blahušův [Blahuš 2000] nabádavý článek zase ukazuje, že ani badatelé ve vědách, které mají blízko k vědám přírodním, na tom nejsou o mnoho lépe. V článku varujeme před (českou) obsesí používat postupy statistické inference vždy, za každou cenu a bez ohledu na typ dat, která analyzujeme.

Z důvodů jisté problematičnosti celého konceptu statistické indukce v sociálních vědách někteří zahraniční autoři radikálně navrhují tyto postupy zcela vypustit ze statistické analýzy. Nejdeme tak daleko, avšak nabádáme k jejich uvážlivé aplikaci. Na to, že v nereflektované aplikaci statistické indukce může být ještě hlubší problém, poukazuje Field, jehož dlouhou citací náš článek uzavíráme.

> „Většina statistik používaná v sociálních vědách je založena na lineárních modelech. Většina výsledků publikovaných v časopisech jsou statisticky signifikantní výsledky. Jelikož se sociálněvědní badatelé většinou učili používat technik, které jsou na

těchto lineárních modelech založené, znamená to, že publikované výsledky jsou ty, které lineární modely využily. Což znamená, že data a vztahy, které mohou být zpracovány na základě nelineárních modelů, jsou povětšinou mylně ignorovány – mylně proto, že na nelineární data byly aplikovány lineární přístupy, takže výsledky badatelům ‚nevyšly'... Je proto možné, že poznatky v některých oblastech vědy se vyvíjejí zkresleně" [Field 2005: 22].

PETR SOUKUP *je vyučujícím na katedře sociologie FSV UK a FSS MU. Soustředí se na výuku a aplikace statistických metod, zejména na multivariační analýzu dat, regresní přístupy a analýzu kategoriálních dat. Z věcného hlediska se zaměřuje na problematiku sociologie vzdělání a environmentální sociologii. V poslední době publikoval článek o lineárních víceúrovňových modelech.*

LADISLAV RABUŠIC *je profesorem brněnské katedry sociologie a v současné době i děkanem Fakulty sociálních studií Masarykovy univerzity. Vyučuje kurzy o metodologii sociálněvědních výzkumů, dále kvantitativní analýzu dat, populační studia a sociologii populačního stárnutí. Badatelsky se soustřeďuje kromě jiného na sociologické aspekty populačních procesů. Dosud publikoval přes sedmdesát statí doma i v zahraničí, je pravidelným přispěvatelem do* Sociologického časopisu / Czech Sociological Review *a* Demografie. *Je autorem monografie* Česká společnost stárne *(1995), editorem publikace* Česká společnost a senioři *(1997) a monografie* Kde ty všechny děti jsou?

Literatura

Anděl, J. 2003. *Statistické metody*. Praha: Matfyzpress.
Anděl, J. 2005. *Základy matematické statistiky*. Praha: Matfyzpress.
Blahuš, P. 2000. „Statistická významnost proti vědecké průkaznosti výsledků výzkumu." *Česká kinantropologie* 4 (2): 53–72.
Blaikie, N. 2003. *Analyzing Quantitative Data*. London: Sage.
Blatná, D. 1996. *Neparametrické metody*. Praha: Vysoká škola ekonomická.
Čermák, V., Vrabec, M. 1998a. *Teorie výběrových šetření* – část 1. Praha: Vysoká škola ekonomická.
Čermák, V., Vrabec, M. 1998b. *Teorie výběrových šetření* – část 2. Praha: Vysoká škola ekonomická.
Čermák, V., Vrabec, M. 1999. *Teorie výběrových šetření* – část 3. Praha: Vysoká škola ekonomická.
Field, A. 2005. *Discovering Statistics Using SPSS*. London: Sage.
Hamplová, D. 2005. „Základní principy víceúrovňových modelů." *SDA Info* 7 (2): 1–2.
Hendl, J. 2004. *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Praha: Portál.
Herzmann J., I. Novák, I. Pecáková. 1995. *Výzkumy veřejného mínění*. Praha: Vysoká škola ekonomická.
Hox, J. J. 1995. *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties.
Hox, J. J. 2002. *Multilevel analysis: techniques and applications*. Mahwah (N.J.): Earlbaum.

Kahounová, J. 2000. *Praktikum k výuce matematické statistiky I.* Praha: Vysoká škola ekonomická.
Knoke, D., G. W. Bohrnstedt, A. P. Mee. 2002. *Statistics for Social Data Analysis.* Belmont (CA): Wadsworth/Thomson Learning.
Kish, L. 1965. *Survey Sampling.* New York: Wiley-Interscience.
Likeš, J., Laga, J. 1978. *Základní statistické tabulky.* Praha: Státní nakladatelství technické literatury.
Norušis, M. 2004. *Advanced Statistical Companion. SPSS.* Upper Saddle River (N.J.): Prentice Hall.
Raudenbush, W. S., A. S. Bryk. 2002. *Hierarchical linear models: applications and data analysis methods.* London: Sage.
Řehák, J., B. Řeháková. 1986. *Analýza kategorizovaných dat.* Praha: Academia.
Soukup, P. 2006. „Proč užívat hierarchické lineární modely." *Sociologický časopis / Czech Sociological Review* 42 (5): 987–1012.
Thompson, S. K. 2002. *Sampling.* New York: Wiley-Interscience.
Vaus, D. A. de. 2002. *Analyzing Social Science Data. 50 Key Problems in Data Analysis.* London: Sage.
Zvára, K., J. Štěpán. 2002. *Pravděpodobnost a matematická statistika.* Praha: Matfyzpress.