

Metoda neúplného zjišťování je pro sociologický výzkum — vzhledem k otázkám, jimiž se obvykle zabývá — nepoužívanějším postupem. Jen v málokterých případech je totiž možné prozkoumat všechny prvky, z nichž se skládá objekt vlastního šetření, ať již z důvodů materiálních, personálních či časových. Vzhledem k tomu je při přípravě většiny sociologických výzkumů jednou z prvních fází volba vhodného výběrového postupu, který zaručí co nejvyšší reprezentativitu informace získané šetřením a při použití adekvátních postupů odhadu tak vytvoří spolehlivý základ pro kvalitní sociologické závěry.

Kniha docenta Václava Čermáka, vydaná v loňském roce, zaplňuje značnou mezeru v původní české literatuře o problematice výběrových šetření. Nepočítáme-li stručnou základní učebnici téhož autora: *Statistika II* [vydanou v roce 1968], je poslední prací tohoto typu dnes již klasická kniha J. Hájka *Teorie pravděpodobnostního výběru s aplikacemi na výběrová šetření*, vydaná před dvaceti lety.

Kniha je určena nejen sociologům, ale i ekonomům, demografům, psychologům či statistikům, kteří nejsou specialisty v oboru výběrových šetření. Metodologům pracujícím v oblasti sociologie umožňuje seznámit se na solidním matematickém základě s klasickou i novější teorií i hlavními možnými aplikacemi výběrových šetření — výkladem od základních pojmů ke složitějším otázkám — a vyžaduje tedy jen základní znalosti pojmů a pouček z matematické statistiky a teorie pravděpodobnosti. Zařazení některých speciálních partií, které jsou novým přínosem k teorii, a formulace některých dosud otevřených problémů ve formě cvičení dávají v některých místech studii charakter monografie.

Kniha se skládá z devíti kapitol:

1. Obecné otázky
2. Prostý náhodný výběr
3. Výběr s nestejnými pravděpodobnostmi
4. Poměrové odhady I
5. Poměrové odhady II
6. Regresní odhady
7. Oblastní výběr
8. Výběr skupin
9. Dvoustupňový výběr

První kapitola ukazuje význam výběrového šetření jako metody neúplného zjišťování. Autor se v úvodu kapitoly snad až příliš stručně zmiňuje o jiných výběrech než pravděpodobnostních, používá zde také termín „reprezentativita“, aniž by jej definoval, což je do jisté míry škoda. Otázka reprezentativity je totiž pro praktickou aplikaci výběrového šetření jednou ze základních a její chápání ve smyslu „pokrytí cílové populace“ je často jen intuitivní. Doporučuji proto čtenáři, aby si při studiu této partie rečené knihy přečetl článek Jana Řeháka v *Sociologickém časopise* 78/5, kde je tato otázka podrobně analyzována.

Stejně tak se domnívám, že by bylo vhodné zmínit se o nesporných styčných bodech ankety a samovýběru, které jsou pro interpretaci dat získaných touto často používanou metodou podstatné a autor se o nich nezmiňuje.

Úvodní kapitola dále vymezuje základní pojmy a používanou terminologii. Za zmínku stojí, že autor důsledně zavádí pojem „estimátor“ místo pojmu „odhad“ pro odhadovou statistiku, čímž odstraňuje známý nedostatek běžně používané terminologie, v níž pojem odhad zahrnuje jak funkci, tak její hodnotu v bodě — vlastně výsledek odhadu.

Po stručné zmínce o ostatních metodách se učebnice soustřeďuje na pravděpodobnostní výběry, které mají i v sociologii velmi dobrou pověst.

Potřebám adekvátního používání metody výběru i odhadu vychází vstříc nejen svým obsahem, ale i tím, že na rozdíl od některých jiných učebnic výběrových šetření autor při výkladu jednotlivých postupů nevychází jen ze schémat výběrů s vrácením, které jsou sice při odvozování elegantnější a dávají i vzorcům jednodušší podobu, i když jen málo odpovídají praktickým potřebám sociologických šetření, ale důsledně používá schémata výběrů bez vrácení. (Výběr s vrácením znamená, že vybraná jednotka je vrácena k ostatním nevybraným před dalším výběrovým krokem a obecně tedy umožňuje další vybrání téže jednotky. Z hlediska teorie pravděpodobnosti tento postup vytváří nezávislé pokusy; odtud jednodušší výpočetní postupy; v sociologické praxi je však jistě ještě častější případ, kdy se vybírá jen z jednotek, které dosud nebyly vybrány.)

Úvodní kapitola dále obsahuje i autorův nástin obecné teorie výběru a odhadu z konečných populací [paragraf 1.3.], v němž jsou dosud spíše časopisecké práce spojeny do uceleného přístupu. Dobře jsou zde zvýrazněny — i když možná spíše pro toho čtenáře, který je již více seznámen s obecnou teorií odhadu — odchylky způsobené konečností populace. I tato partie dokládá autorovo přizpůsobení potřebám praxe, v níž je konečný základní soubor nejčastější. Původní prací je i druhá „poznámka k intervalovým odhadům“, která se zabývá stanovením přípustné chyby v případě odhadu s mírně šikmým rozložením, což se v praxi vyskytuje zejména u rozsáhlejších výběrů.

Druhá kapitola obsahuje teorii prostého náhodného výběru v obvyklém rozsahu i uspořádání. Prostý náhodný výběr, který je definován jako takové výběrové schéma, při němž mají všechny stejně veliké výběrové soubory stejnou pravděpodobnost vytvoření, má nesporně svůj význam pro základní pochopení principů pravděpodobnostních výběrů, v sociologických výzkumech však není (snad s výjimkou průmyslové a podnikové sociologie) příliš používán. Jeho aplikace totiž předpokládá existenci úplného centrálního seznamu všech jednotek, což v praxi — až na výše uvedené případy — vždy neexistuje, a navíc tato metoda neumožňuje využít některé obvykle předem známé informace o základním souboru ke zvýšení kvality získaných odhadů. Proto se v praxi používají složitější techniky, u nichž stačí i jinak uspořádaný seznam jednotek a navíc umožňují výběr zefektivnit.

Z obvyklého rozsahu se poněkud vymyká paragraf 2.6., který se zabývá problematikou odhadů založených na různých prvcích výběrového souboru při výběru s vrácením. Autor ukazuje, že jím doporučené pravidlo k zastavení výběrové procedury vede ke zjednodušení odhadových vzorců.

V kapitole třetí je vyložen výběr s nestejnými pravděpodobnostmi, jehož motivací je (oproti prostému náhodnému výběru) předem známý nestejný význam jednotlivých prvků základního souboru z hlediska některé proměnné. Této apriorní informace se používá ke zvýšení vydatnosti odhadů ve srovnání s odhady prostého náhodného výběru. Toto zvýšení, nazývané obvykle zisk na vydatnosti, je navíc autorem vyjádřeno i novým způsobem zavedením nové míry kolem úměrové regrese (s. 140).

Čtvrtá a pátá kapitola se zabývá problematikou poměrových odhadů, tedy odhadů poměrových veličin, odhadů podílu dvou proměnných. Tato otázka je typická třeba pro demografické či zemědělské výzkumy. Obtížnost stanovení odhadových statistik pro proměnné tohoto typu plyne z toho, že se jedná vlastně o odhad složené náhodné veličiny. Kapitola čtvrtá zkoumá — vedle výkladu základní používané teorie — zejména podmínky, za nichž je použití poměrového odhadu proti prostému výhodné. Obecně totiž neplatí, že je poměrový odhad lepší než prostý náhodný a analýza této otázky je pro praktické rozhodnutí o výběrovém postupu velmi významná.

Pátá kapitola se pak zabývá některými speciálnějšími a novějšími postupy poměrových odhadů. V první řadě je to otázka vychýlenosti poměrových odhadů a její redukce, což je jeden z často zkoumaných problémů výběrových šetření. Běžně používané metody poměrových odhadů, o nichž pojednává kapitola čtvrtá, jsou totiž obecně vychýlené, to znamená, že střední hodnoty používaných odhadových statistik se liší od střední hodnoty odhadované veličiny. To může v praxi vést k nepřijemnostem zejména u stratifikovaných odhadů s více oblastmi, kde jsou rozsahy výběrů v jednotlivých strtech relativně malé. Autor se zabývá jak otázkou odhadu velikosti vychýlení a jeho následné redukce, tak postupy vedoucími ke konstrukci nevychýlených odhadů. Analýza podmínek výhodnosti tohoto postupu je spolu s rozбором přínosu druhé pomocné proměnné do vícenásobných poměrových odhadů v paragrafu 5.2.1. původním přínosem k teorii výběrových šetření.

Ukazuje se, že poměrové odhady jsou nevhodné v těch případech, kdy regrese obou odhadovaných veličin ani neprochází počátkem, ani není v jeho blízkosti. V těchto případech je vhodnější použít metodu regresních odhadů, která je obsahem výkladu šesté kapitoly. Regresní odhady využívají apriorní informace o základním souboru podobně jako poměrové odhady zavedením pomocné proměnné, jsou však v jistém smyslu jejich zobecněním. (Odhadneme-li regresní koeficient poměrem výběrových průměrů, dostaneme jeden z používaných poměrových odhadů.) Lze říci, že regresní odhady jsou v zásadě vhodné pro případy, pro něž jsou používány odhady poměrové, lze je však použít i tehdy, kdy poměrové odhady nejsou vhodné. Autorovým přínosem k teorii je v této kapitole určení střední čtvercové chyby u regresního odhadu s koeficientem regrese odhadnutým z výběru.

Poslední tři kapitoly knihy se zabývají otázkami složitějších uspořádání výběru, která předem svou konstrukcí vylučují vznik některých výběrových souborů, především těch, jež by vedly k silně vychýleným odhadovým statistikám.

Základní soubor je u těchto metod nejprve rozdělen na několik částí; ve všech těchto částech se buď vybírá určitý počet jednotek, což vede k oblastnímu, neboli stratifikovanému výběru, nebo je další výběr prováděn jen v některých částech, což je princip dvoustupňového, popřípadě vícestupňového výběru. Jestliže je ve vybraných částech prováděno úplné šetření, a tedy zkoumány všechny jednotky, jedná se o skupinový výběr.

Stratifikovaný či oblastní výběr, jehož teorie je vyložena v kapitole 7, je další možnost, jak vedle poměrových či regresních odhadů efektivně využít známé informace o základním souboru. Zde je pomocná proměnná využita k vymezení oblastí (neboli strat), v nichž jsou poté prováděny nezávislé náhodné výběry. Velikost výběrů v jednotlivých oblastech může být volena buď proporcionálně k jejich velikosti, nebo — v případě dalších informací — je možné použít takzvané optimální rozvržení výběru do oblastí. Výhoda tohoto postupu je ve snížení rozptylu odhadů; toto snížení je přitom tím větší, čím větší část variability je tvořena rozptyly mezi oblastmi, a nikoli rozptyly uvnitř jednotlivých oblastí, to znamená že optimální rozvržení odhadu je nejvýhodnější tehdy, jestliže se základní soubor skládá z částí navzájem sice velmi odlišných, vnitřně však velmi homogenních. Nevýhodou optimálního rozvržení je zejména požadavek znalosti rozptylů v jednotlivých oblastech, dále pak značná komplikace při odhadu více proměnných, kdy na základě každé z nich dostáváme obecně různé optimální velikosti výběru v jednotlivých oblastech. Touto, pro praxi velmi závažnou otázkou (málokdy je odhadována jen jedna proměnná a nelze ani zaručit podobnou variabilitu proměnných v jednotlivých oblastech, aby bylo možné rozumně vybrat kompromisní velikost, alespoň přibližně vhodnou pro všechny proměnné simultánně) se autor — pravděpodobně vzhledem k převážně didaktickému poslání knihy — bohužel nezabývá.

V případě, že je informace o rozptylech uvnitř oblastí nepřesná nebo chybí, je použití optimálního rozvržení nevhodné a je lepší použít proporcionální stratifikaci. Pro volbu vhodné stratifikační proměnné dále platí, že efekt stratifikace, čili snížení rozptylu odhadů, je tím větší, čím je větší koeficient korelace mezi odhadovanou a stratifikační proměnnou. Výhodou oblastního náhodného výběru je vedle stratifikačního efektu i to, že nepožaduje existenci centrálního seznamu jednotek, ale jen seznamů v jednotlivých oblastech.

Osmá kapitola pojednává o skupinovém náhodném výběru, který se s výhodou používá tehdy, kdy neexistuje evidence elementárních jednotek, ale existuje evidence jejich určitých seskupení. Typickým příkladem pro sociologickou praxi je výběr domů v městské lokalitě nebo několika dílen v závodě, v nichž jsou pak všichni obyvatelé nebo pracovníci podrobeni šetření. Metoda je výhodná i ekonomicky, neboť umožňuje soustředění sil do jednoho místa i času.

Poslední kapitola se zabývá otázkami dvoustupňového výběru, kdy je nejprve proveden náhodný výběr několika oblastí, do nichž byl základní soubor rozdělen, a v nich teprve náhodný výběr vlastních prvků. Tento postup je sice ekonomicky velmi výhodný u výběru z velkých základních souborů (například z celé populace ČSSR), získané odhady jsou však zatíženy poněkud větší chybou.

Složitější, vícestupňové postupy organizace výběrových šetření pak mohou různě kombinovat metody uvedené v posledních třech kapitolách, v praxi našich výzkumů se však vyskytují zřídka.

V dodatku, který je určen spíše pro matematiky, je vyložena teorie momentů a dále moderní teorie tzv. k -statistik, kterou autor použil v předchozím textu ke zkoumání vlastností poměrových odhadů i k přesnějšímu určení střední hodnoty a čtvercové chyby u odhadů regresních.

Závěrem lze shrnout, že kniha Václava Čermáka je nesporným přínosem pro metodologii sociologických výzkumů. Jestliže se v citovaném článku Jana Řeháka říká, že „Výběrové šetření můžeme vyhodnotit jen tehdy, jestliže vytvoříme matematický model postupu, který určuje všechny složky výběrového uspořádání, umožňuje určení výsledných výpočetních formulí a výpočet pravděpodobností pro výběr každé jednotky na všech stupních“, pak právě recenzovaná učebnice výběrových šetření umožňuje tomu, kdo ji prostuduje, takovéto exaktní modely vytvářet ve své praxi.