

Je-li vydána první monografie na nové téma, znamená to, že se zkoncipoval nový specifický směr vědeckého bádání, že se nahromadilo již tolik poznatků v řešení určitého problému a nový přístup k nějakému problému se již tak vyvinul, že nové, v časopisecké literatuře většinou rozptýlené výsledky potřebují shrnující přehled. Monografie (a často to bývá celá řada paralelních monografií) uzavírají vždy určitou ucelenou a relativně uzavřenou etapu vývoje poznání v daném směru. A zároveň svou uceleností přinášejí nový pohled na problém, a tudíž podněty k dalšímu rozvoji. Proto recenze takové práce představuje zároveň pohled na úspěchy či neúspěchy, použitelnost či nepoužitelnost celého směru a jeho dosavadních výsledků i případnou úvahu o dalších směrech žádoucího vývoje.

Agrestiho knížka představuje ucelený pohled na ten směr v metodologii analýzy dat, který vyvíjí metody a zpracovatelské techniky pro ordinální proměnné, směr, který zaznamenal v posledních deseti letech velmi značné oživení zájmu mezi statistiky a který přinesl nové zajímavé a prakticky důležité výsledky. Po dlouhou dobu se ordinální data zpracovávala buď jako nominální (tím, že se prostě zapomnělo na uspořádání kategorií), nebo jako kardinální (tím, že byla přijata nějaká vhodná, většinou ekvidistantní kvantifikace). Z čistě ordinálních technik byla využívána jen pořadová korelace (Spearmanův nebo Kendallův koeficient), modifikovaná pro kontingenční tabulky a (vzhledem k porušení předpokladu spojitosti velmi nepřesně a chybně) Kolmogorov-Smirnovova statistika pro testy dobré shody nebo dvouvýběrový problém. První vlna oživení zájmu přichází v šedesátých letech — potřeba měření asociace ordinálních znaků, vstup ordinálních dat do modelů kauzálních sítí a mnohorozměrné analýzy se projevila především ve společenských vědách. V posledních letech se nový zájem přenáší na pole matematické statistiky, na němž už vznikají ucelené metodologie, v anglicky publikované literatuře především log-lineární a logitové modely, u nás pak modely kvadratických forem distribučních funkcí.

A. Agresti již v úvodu shrnuje čtyři důvody pro tento nový silný zájem, které současně považuje za výhody oproti nominálním přístupům:

1. Ordinální metody mají větší sílu v odhalení specifických alternativ k nulové hypotéze.
2. Ordinální data jsou popisována podobně jako metodika regresní analýzy a analýzy rozptylu pro spojitá číselná data.
3. Ordinální analýza může využít širší možnosti ve formulaci modelů, které jsou úspornější a mají jednodušší interpretaci než standardní modely pro nominální proměnné.
4. Zajímavé ordinální modely mohou být aplikovány v situacích, kde standardní nominální modely jsou triviální nebo mají příliš mnoho parametrů pro testování shody.

Tyto čtyři vlastnosti jsou bezesporně velmi závažné a shrnují statistický pohled na metodiku analýzy ordinálních dat, domnívám se však, že je to pohled jednostranný a nezahrnuje to, co je z hlediska aplikačního předmětného oboru nejdůležitější: ordinální znaky jsou samostatnou metodologickou entitou, neboť kategorizace s uspořádáním tříd je vhodným modelem pro velmi širokou třídu vlastností, jež jsou pomocí něho vhodně operacionalizovány a vyjádřeny znakem. Ať již takto přímo vyjádřený, nebo jen pocívaný a latentně a intenzivně působící, je tento důvod vlastním motivem rozvoje metodiky. Uspořádanost stavů vlastnosti je součástí jejího významu, kterou nelze trvale pomíjet ani libovolně kvantitativně ohodnocovat (a tím vlastnost v průběhu operacionalizace zkreslovat). To je také důvod intenzivního zájmu společenských věd a jejich publikačních periodik o tuto problematiku, neboť právě v nich se ordinální data vyskytují velmi často. Rozbor těchto aspektů, který má základ v gnozeologických a ontologických úvahách, nemůžeme ve statistické monografii očekávat. Jejich absence však vede například k dalším jednostrannostem: autor de facto ztotožňuje ordinální proměnnou s intervalovou, předpokládá, že každé ordinální proměnné je podložena spojitá kvantitativní proměnná. Definuje (kap. 1): „Kategorizovaná proměnná je nazývána „ordinální“ místo „intervalovou“, když existuje jasné pořadí kategorií, ale absolutní vzdálenosti mezi nimi nejsou známy.“ Odhlédneme-li od nepřesnosti v požadavku pro absolutní vzdálenosti, který odporuje předpokladu intervalového typu (u toho lze pracovat jen s poměrnými vzdálenostmi vzhledem k invarianci v měřítku), vyvstává tu několik otázek. Především uspořáda-

nost nemusí vycházet z kvantit, je tu chudší relační struktura (která ovšem může také vzniknout absencí číselné informace, i když nejen intervalového typu) na množině stavů vlastností. A dále ordinální proměnná může být diskrétní svojí podstatou, její třídy jsou samostatné, ohraničené a jsou uspořádány. Předpoklad číselné určující proměnné určuje uspořádání všech jednotek na nějaké spojitě škále, zatímco diskrétnost předpokládá pouze uspořádání kategorií, nikoli jednotlivců v nich. Pojem ordinality metod ovšem v obou případech od kvantifikace odhlíží: metoda je ordinálního charakteru, jestliže její výsledky jsou nezávislé na jakékoli kvantifikaci a reflektují pouze uspořádání kategorií, přičemž se při změně pořadí mění i výsledky.

V současné době lze přístupy k rozvíjení metodiky rozdělit na tři hlavní směry:

- I. Přiřazení neparametrických skóreů za předpokladu spojitých proměnných.
- II. Využití distribučních funkcí pro případ diskrétních proměnných.
- III. Přiřazení číselných skóreů kategoriím znaku a přechod k metodám analýzy kardinálních dat.

V knize se autor zabývá především log-lineárními modely, které patří do III, logitovými modely, které patří částečně do II (závislá proměnná) a částečně do III (explanační proměnné), metody typu I se v knize vyskytují v souvislosti s měřením asociace a v porovnání závislých distribucí.

První čtyři kapitoly obsahují základní pojmy a pomocný aparát pro vlastní látku. Je tu shrnuta metodika log-lineárních modelů pro standardní nominální analýzu v tabulkách $R \times C$ a $R \times C \times L$, uvádějí se modely marginální i parciální asociace, interakční struktura, hierarchické modely. Jsou tu také popsány numerické algoritmy pro odhad parametrů. Jedním z důležitých pojmů, který se u nás neujal (přes svoji neoddiskutovatelnou teoretickou i praktickou závažnost) je „odds ratio“, které je jediným dnes známým koeficientem asociace v tabulkách 2×2 , jehož hodnota je necitlivá na změnu v poměrech marginálních četností a hraje ve formulaci log-lineárních modelů nezastupitelnou roli. K tomu terminologická poznámka: „odds ratio“ je možno přeložit buď jako „poměr šancí“ nebo jako „součinový poměr“. Vzhledem k tomu, že první termín má v některých sociologických oborech (například sociální mobilita) specifické předmětné konotace, které mohou být interpretačně zavádějící, je vhodnější přijmout neutrálnější druhou možnost, která podobně nebezpečí neobsahuje (takovou preferenci navrhuji i pro případ, že by v ostatních vědách či ve statistice převážil termín první).

První čtyři kapitoly tvoří odrazový můstek k vlastnímu tématu, které začíná kapitolou pátou, *Log-lineární modely pro ordinální proměnné*. Tyto modely vycházejí z přiřazených skóreů kategoriím a z lineárních vztahů mezi logaritmy očekávaných četností. Pro dvourozměrnou tabulku $R \times C$ je uveden model asociace, v níž jsou obě proměnné ordinální a model, v němž je jedna z proměnných nominální (model řádkových, resp. sloupcových efektů). Vícerozměrné tabulky jsou reprezentovány případem trojrozměrným, $R \times C \times L$, který vyjadřuje již širší možnosti formulace závislostních vazeb a i složitější kombinace ordinálních a nominálních proměnných. Jsou to především modely stejnoměrné interakce, homogenní stejnoměrné interakce, řádkových, resp. sloupcových efektů v různých obměnách. Využití přiřazených skóreů v modelových rovnicích zvyšuje ovšem typ proměnných na kardinální. Výsledky logaritmicko-lineárních modelů jsou na zvolené kvantifikaci závislé, a tudíž nesplňují podmínky ordinality metod. Jsou příkladem pragmatického přístupu statistiků k řešení problému: využití vhodné a matematicky zpracovatelné teorie dává aplikovatelné, pochopitelné a názorné, i když ne zcela adekvátní a vyčerpávající řešení. V bilanci „pro a proti“ však modely získávají mnoho bodů pro přijetí v praxi (byť by měly být v budoucnu překonány a využití bylo jen dočasné). Ve většině případů se nevyskytují tak divoké interakční struktury, aby přijetí či odmítnutí modelů vedlo k velkému a interpretačně nebezpečnému zkreslení pohledu výzkumníka. Z textu není tak zřejmé, zda aplikace ekvidistantních skóreů nevede snad k žádoucím ordinálním vlastnostem, jak to naznačují vlastnosti součinových poměrů, tj. nevede-li k formulaci modelů, jež se bez lineárních vazeb ke skóreům obejdou. Problém rozporu mezi nutností kvantifikovat a požadavkem ordinality autor uvádí, ale neřeší. Nejistotám ohledně přijatých skóreů se můžeme částečně vyhnout tím, že analýzu opakujeme několikrát pro značně různé kvantifikace kategorií. V případě, že se sociologické interpretace získaných výsledků od sebe neliší, posiluje se naše přesvědčení o platnosti výsledků a o jejich nezávislosti na možném omylu při určování skóreů.

Logaritmicko-lineární modely nerozlišují závislé a explanační (nezávislé) proměnné, směr působení je součástí meritorního modelu. Výsledky, které nabízí statistika, popisují statistické vazby, konstrukce návazných kauzálních modelů je záležitostí

sociologie. (Takováto striktní dělba práce a odpovědnosti však v praxi neplatí a konstrukce modelů je neustálou interakcí statistického a sociologického myšlení, přístupu a poznatkových fondů).

Další dvě kapitoly se zabývají situací, v níž je apriori známo, že jedna z proměnných je závislá, že zkoumáme efekty působení ostatních proměnných na ni. Tento případ je řešen pomocí logitových modelů.

Šestá kapitola má opět přípravný charakter, uvádí základní pojmy, z nichž nejdůležitější jsou: logistická regrese, logistická křivka a především pojem logitu pro relativní četnost (dichotomickou proměnnou), logitové báze pro vícehodnotový nominální znak a modifikaci pojmu pro ordinální případ (kumulované logity, přírůstkové logity, logity sousedních kategorií).

Sedmá kapitola, *Logitové modely pro ordinální proměnnou*, řeší případ dvourozměrných a třírozměrných tabulek s asymetrickou asociací. Problém rozpornosti kvantifikace a ordinality se projeví jen u explanačních proměnných. Čtenář se zde seznámí s bohatou varietou situací a modelů.

Kapitola 8, *Jiné modely pro ordinální proměnné*, je stručný přehled dalších přístupů k řešení ordinálního problému, které se vyskytují v anglicky publikované literatuře. Jsou to především logaritmicko-multiplikační modely, u nichž vycházíme z předpokladu, že kvantifikační skóry jsou neznámými konstantami, jež se v rámci modelu odhalují. Tento aspekt však vnáší do problému nové komplikace. Výsledné skóry jsou závislé na četnostech v tabulce a z nich odvozené, a tudíž neodpovídají postavení proměnné v celém, širokém systému ostatních proměnných a vztahů k nim. Kvantifikace tu znamená charakteristiku vzájemnosti, a to bez ohledu na uspořádanost kategorií, danou věcným a na datech nezávislým opodstatněním, která určuje obsah ordinálního znaku. Modely jsou formulovány zcela stejně pro nominální proměnné, uspořádanost kategorií nevyužívají. Jsou zcela analogické metodě kanonické korelace v kontingenčních tabulkách a dávají také obdobné výsledky. Logaritmicko-multiplikační modely mají přímou souvislost s metodou LINDA - A, což také naznačuje faktorovou nebo korespondenční interpretaci získaných hodnot.

Jiný přístup ukazuje Plackettova třída rozložení se stejnouměrnou globální asociací, modely kumulačního efektu, modely průměrné odpovědi.

V deváté kapitole, *Míry asociace pro ordinální proměnné*, se setkáme s běžnými tradičními měrami (založenými na přístupu I) i s měrami zavedenými autorem. V následující kapitole se čtenář seznámí s postupy statistické inference pro koeficienty asociace, včetně parciálních. Kapitola o čtvercových tabulkách obsahuje několik modelů pro quasisymetrii tabulky a test pro rovnost marginálních distribucí, není však založena na jednotném metodicko-teoretickém přístupu. Poslední kapitola přináší souhrnné porovnání log-lineárních, logitových modelů a koeficientů asociace, jsou tu stručně zmíněny některé metodologické aspekty, které s problémem souvisejí: nominální vs. ordinální data, měření asociace vs. budování modelů, logitové vs. log-lineární modely.

Dodatky obsahují technické výsledky, nutné pro realizaci metod na počítačích a pro odvozování vlastních a výpočetních rovnic: metoda vážených nejmenších čtverců, metoda maximálních věrohodností, delta metoda a přehled počítačových programů, které obsahují možnosti aplikace modelů uvážených v knize.

Každá kapitola je zakončena poznámkami a doplňky, které obsahují další vlastnosti, dodatečné informace, paralelní metody, ale i zajímavá data a cvičení. Většina modelů knihy je ilustrována na příkladech. Autor předpokládá, že čtenář je obeznámen s pojmy testování hypotéz, ví co je to asociace a k čemu se ve vědě používá. Proto se soustřeďuje na statistickou stránku věci a i výsledky analýz na datech příkladů interpretuje na statistické rovině, nikoli předmětné. Z tohoto hlediska (které zřejmě je i cílem autorovým) je třeba se na knihu dívat.

A. Agresti pečlivě shrnuje literaturu z anglo-americké provenience, soustředil se na dva modely, které jsou v praxi nejčastěji používané a v teorii nejdále rozpracované a napsal stručný, dobře pochopitelný přehled statistické metodiky analýzy ordinálních dat. V tom je také velký přínos knihy pro praktické statistiky i pro výzkumníky, kteří mají aspirace na vlastní hlubší analýzu dat pomocí log-lineárních a logitových modelů a na hlubší a samostatné porozumění těmto modelům a jejich významu. Kniha je shrnutím dosavadních výsledků, které se budou ještě dále rozvíjet a prohlubovat, jistě se prosadí i jiné přístupy a metodika se rozšíří. Postupy, které A. Agresti v knize shrnul, však již tvoří dostatečně obsáhlý systém, aby se stal v blízké budoucnosti nepostradatelným pro každého výzkumníka. To platí v sociologii obzvlášť, vzhledem k roli, kterou ordinální data v našich výzkumných souborech hrají. Monografické zpracování látky, pro sociologii tak důležité, dává také předpoklad k tomu, aby se metodika dostala do vysokoškolských výukových plánů výzkumné a analytické metodologie pro adepty naší sociologie.

Jan Řehák