

Posudek na diplomovou práci

Zdeněk Šulc

Metody analýzy vícerozměrných kontingenčních tabulek

Ačkoliv jsou datové matice obsahující značný podíl či dokonce výlučně kategoriální proměnné běžným výsledkem výzkumů prováděných dotazováním obyvatel, relativně mladé metody vícerozměrné analýzy kategoriálních dat nebývají používány příliš často. Analýzy se omezují na dvourozměrné kontingenční tabulky, které však mohou vést až k nesprávné představě o vztazích ve skupině většího počtu proměnných. Téma diplomové práce proto směřuje právě k analýze vícerozměrných kontingenčních tabulek.

Autor zvolil dvě metody, a sice loglineární modelování a korespondenční analýzu. Cíl práce je konkrétně specifikován v jejím úvodu. Autor hodlá metody prostudovat, aplikovat na reálná data, vyhodnotit jejich možnosti a konfrontovat zjištění. Zdůrazňuje rovněž, že jeho práce má přispět ke zkoumání použitelnosti metod při řešení praktických úloh, což osobně považuji u podobných prací za největší jejich přínos.

Lze říci, že práci tvoří tři části. V první z nich (čtyři kapitoly) autor rekapituluje vymezení kategoriálních proměnných, pojetí vícerozměrných kontingenčních tabulek a nejdůležitější nespojitá pravděpodobnostní rozdělení, která se používají jako výběrová rozdělení pro četnosti v tabulkách. Ve druhé se věnuje základní charakteristice obou pojednávaných metod, tedy korespondenční analýzy a loglineárního modelování. Třetí část práce je aplikační; autor formuluje a řeší úlohy s využitím reálných dat pocházejících z výzkumu zabývajících se mj. názory na podnikatelské prostředí v ČR.

Struktura práce je logická. Lépe snad bylo možné do pojednávané problematiky zasadit téma 4. kapitoly, kde není souvislost s vícerozměrnou analýzou úplně zřejmá. Za nejméně zdařilou lze vůbec považovat první třetinu práce. Některé zde použité formulace jsou přinejmenším neobratné, objevují se ale i nepřesnosti. Uvedu zde jen několik příkladů:

- str. 5: v oboru hodnot náhodné veličiny s binomickým rozdělením chybí nula;
- str. 6: pro binomické rozdělení s parametry $n = 50$ a $\pi = 0,9$ je aproximace normálním rozdělením nevyhovující (n by mělo být větší);
- str. 13: neúplně jasný význam výrazu (3.3);
- str. 14: použité formulace;
- str. 16: nerozumím formulaci, že „koncept podmíněné nezávislosti hraje roli v např. ve faktorové analýze“;
- str. 17: při uvedení parametru rozdělení testového kritéria je zřejmě překlep v předložce, což poněkud mění význam řečeného;
- str. 43: nejedná se zde o *testová* kritéria;

atd.

Naopak lze ocenit, že závěr práce není pojednán jako pouhá rekapitulace jejích částí, jak tomu někdy bývá, ale že autor skutečně formuluje zjištění, k nimž v práci dospěl.

Nezaznamenala jsem žádné gramatické či stylistické chyby a jen minimum překlepů, použité vzorce jsou v náležitém formátu a jsou číslovány, stejně tak tabulky a grafy zařazené do textu. Rozměrnější počítačové výstupy a syntaxe použitého systému (SPSS), kterou autor považoval za užitečnou, jsou pak uvedeny v přílohách práce. Seznam literatury a použitých pramenů obsahuje přiměřený počet položek. Rozsah textu je přiměřený diplomové práci.

Pro diskusi u obhajoby práce navrhuji následující témata:

- 1) Problém analýzy kategoriálních dat známý jako „overdispersion“ a možnosti jeho řešení.
- 2) Možnosti hodnocení kvality loglineárního modelu

Některé formulační nepřesnosti, které se v práci objevují, jsou však podle mého názoru kompenzovány náročností tématu a také s aplikační částí práce jsem víceméně spokojena. Předloženou práci proto doporučuji k obhajobě před příslušnou komisí, které zároveň navrhuji hodnotit ji jako *velmi dobrou*.

V Praze 4. 6. 2012

Doc. Ing. Iva Pecáková, CSc.
vedoucí diplomové práce