

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

4/2015

ročník/volume 25

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 12

Typ článku/Type of article: recenzia publikácie/review of publication

Strany/Pages: 110 – 111

Dátum vydania/Publication date: 15. október 2015/October 15, 2015



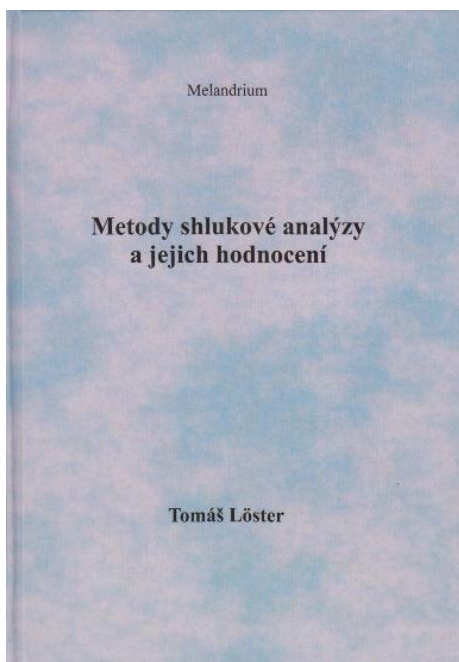
Recenzia publikácie/Review of Publication

Tomáš Löster:
METODY SHLUKOVÉ ANALÝZY A JEJICH HODNOCENÍ

METHODS OF CLUSTER ANALYSIS AND THEIR EVALUATION

Melandrium, Slaný 2014, 132 s.

ISBN 978-80-86175-88-1



Publikace s názvem *Metody shlukové analýzy a jejich hodnocení* se zaměřuje na skupinu metod analýzy dat, která je velmi užitečným nástrojem jak ve vědeckém výzkumu, tak při řešení praktických úloh, zejména ekonomicky zaměřených. Autor je odborným asistentem na Fakultě informatiky a statistiky Vysoké školy ekonomické v Praze a problematice shlukové analýzy se věnuje dlouhodobě.

Na rozdíl od mnohých publikací, které popisují zejména základní postupy shlukování objektů, které jsou charakterizovány hodnotami kvantitativních proměnných, je monografie Tomáše Löstra rozšířena dvěma směry. V prvním z nich jsou zohledněny také jiné než kvantitativní proměnné. Pozornost je věnována jednak datovým souborům pouze s kvalitativními proměnnými, jednak souborům dat s proměnnými různých typů. Druhý směr rozšíření se týká shrnutí poznatků o hodnocení výsledných skupin objektů (shluků), které lze využít jednak k výběru co možná nejlepšího přiřazení objektů do shluků (z řešení získaných různými metodami), jednak ke stanovení optimálního počtu shluků.

Publikace sestává z šesti kapitol. První se zaměřuje na měření podobnosti objektů a shluků a na vybrané metody shlukové analýzy. Další čtyři kapitoly jsou věnovány hodnocení získaných řešení. Poslední (šestá) kapitola doplňuje předchozí teoretický výklad aplikací jedné z uvedených metod na data z oblasti ekonomické statistiky stanovením optimálního počtu shluků na základě vybraných kritérií a charakteristikou jednotlivých shluků.

Autor popisuje následující postupy shlukové analýzy: hierarchické metody (nejbližšího souseda, nejvzdálenějšího souseda, průměrné vzdálenosti, centroidní, mediánovou a Wardovu), metody pro pevné k -shlukování (např. k -průměrů a k -medoidů), metody pro fuzzy shlukování (fuzzy k -průměrů a algoritmus FANNY) a metody určené pro datové soubory s kvalitativními proměnnými a proměnnými různých typů (algoritmy MONA, ROCK, COOLCAT a dvoukrokovou shlukovou analýzu).

Hodnocení získaných řešení je rozčleněno na porovnání výsledků shlukování se známým zařazením do skupin, hodnocení disjunktního shlukování, fuzzy shlukování a shlukování v případě kvalitativních proměnných a proměnných různých typů. Je

popsáno více než 50 různých koeficientů, indexů a statistik. Každá kapitola je ukončena shrnutím, kde je mj. uvedeno, v jakém softwaru jsou pojednávány koeficienty implementovány.

Praktická aplikace spočívá v analýze založené na dvanácti makroekonomických ukazatelích sledovaných u 27 zemí Evropské unie v roce 2011 (zdrojem je databáze Eurostatu). Použita byla Wardova hierarchická metoda. Na základě tří vybraných hodnotících koeficientů byly jako optimální stanoveny čtyři shluky. I když podle dendrogramu, který znázorňuje postupné shlukování, by měl být výsledkem jeden větší shluk zemí a tři méně zastoupené shluky, popsány jsou tři výrazné skupiny tvořící první shluk a jedna skupina tvořená zbylými státy. Zhodnocení těchto skupin zemí spočívá v uvedení základních statistických charakteristik pro šest základních ukazatelů a v porovnání skupinových průměrů s průměrnými hodnotami za všechny země EU.

Čtenáři se v kapitole zaměřené na aplikaci metod mohou seznámit také s vybranými výstupy tří statistických programových systémů. I když by pro samotné analýzy bylo možno využít pouze systém, v němž jsou implementovány koeficienty pro stanovení vhodného počtu shluků (SYSTAT), z důvodu vhodné prezentace dendrogramu a snadného uspořádání charakteristik do tabulek autor využil ještě dva další programové systémy, kterými jsou STATISTICA a IBM SPSS.

Problematika shlukové analýzy je velmi obsáhlá, tudíž je zřejmé, že v útlé publikaci jde skutečně o výběr metod a postupů a o jednu z široké škály mnohých jiných praktických aplikací. Důraz je kladen zejména na zhodnocení výsledného rozdělení objektů do skupin. Naopak, publikace se nezaměřuje na přípravu dat, například na transformaci ukazatelů, aby na shlukování neměly vliv různé měrné jednotky ukazatelů.

V současné době je připravována druhá verze této publikace o shlukové analýze, která by měla vyjít na jaře příštího roku. Pozornost by měla být věnována například zhodnocení použitelnosti vybraných koeficientů pro stanovení počtu shluků. Vhodné by bylo rozšíření ukázek aplikací o analýzu datových souborů s kvalitativními proměnnými, i když možnosti programových systémů v této oblasti stále zaostávají a nepřinášejí žádný výrazný posun. Naopak, v ekonomické oblasti se ukazatele tohoto typu vyskytují, příkladem je šetření domácností EU-SILC.

Prof. Ing. HANA ŘEZANKOVÁ, CSc.

Autorka je zástupkyňou vedúceho Katedry štatistiky a pravdepodobnosti na Vysokej škole ekonomickej v Prahe.