

Oponentský posudek bakalářské práce

Klára Juráňová: Statistická analýza kompozičních dat pomocí knihovny „compositions“ softwaru R

Předloženou bakalářskou práci Kláry Juráňové považuji za velmi zdařilou. Coby oponent jsem se snažil poukázat na některé nedostatky, které však nepovažuji za příliš zásadní či snižující kvalitu práce.

První kapitole (teoretická část) takřka není co vytknout, snad jen místy příliš zjevná podobnost textu s citovanými zdroji (např. odstavec *Porovnání jednotlivých typů souřadnic* odpovídá poznámce 2.2 na str. 44 citované monografie [3]). Nebylo zde patrně dodrženo zásady, že text se píše až poté, co je zdrojová literatura prostudována a bezpečně uschována v šuplíku.

Nejpodstatnější připomínkou k druhé (prakticky orientované) kapitole je absence bližšího popisu způsobu sběru dat o spokojenosti obyvatelstva. I když používáme data pouze k ilustraci, měli bychom k nim chovat respekt a více se jim věnovat (kdo, proč a jak je sbírá). Vzhledem k orientaci práce na výpočetní software chybí přiložené CD s kódy, které by si zvědavý čtenář jistě rád vyzkoušel. Několik nepodstatných připomínek či námětů k druhé kapitole uvádím níže.

Práci navrhuji k obhajobě a doporučuji hodnotit stupněm A.

- s. 37: Z textu v prvním odstavci se poněkud vytrácí sdělení, čeho se vlastně zmiňovaná variabilita týká. Píše se např. „variabilita se projevuje spíše mezi jednotlivými formami vztahů“. Podle mého by bylo lépe napsat např. „Podíl domácností *single* a *other* je napříč státy značně variabilní.“
- s. 37: $clr_i(\mathbf{x})$, $clr_j(\mathbf{x})$, indexy by měly být spíše u $\mathbf{x}$.
- s. 38: Mělo by se poznamenat, že na obrázku 2.10 je sice y -ová osa v logaritmickém měřítku, popis hodnot se však týká samotných podílů, nikoliv jejich logaritmů. Popis obrázku „boxploty logaritmů podílů“ je tak poněkud matoucí.
- s. 39: Mělo by být explicitně uvedeno, že symbol $\mathcal{F}_p$ značí p -kvantil Fisherova rozdělení (podobně u vzorce na str. 40).
- s. 39: Nemá být jasné, proč se mluví o elipsách a co v tomto případě znamená poloměr. Čtenář, který není s problematikou dopředu blíže obeznámen, bude mít s pochopením problémy.
- s. 40: Normalitu netestujeme „abychom splnili předpoklady“, ale „abychom ověřili splnění předpokladů“.
- s. 40: Poslední věta by spíše měla znít „Protože jsou p -hodnoty větší než daná hladina významnosti, žádná z nulových hypotéz nebyla zamítnuta.“
- s. 42: Co znamená „implicitní hodnota“? Je tím míněno „defaultně nastavená“? (též na str. 43)
- s. 43: Spojení „tradiční stát“, za který je později označeno např. Chorvatsko, je poněkud nešťastně zvolené (většinou se užívá ve významu „dlouho existující stát“).

- s. 43: Nerozumím tvrzení „nejsou z hlediska podílů mezi složkami rozdíly v tom, jestli mají jednotlivé typy domácností děti či ne“. Zkuste ho formulovat lépe (viz připomínka ke straně 37).
- s. 43: Život „single“ nemusí být „způsoben“ jen tím, že „člověk dává přednost kariéře před rodinným životem“. Je to podstatně složitější jev. Právě takovéto „zkratky“ nedělají statistikům dobré jméno.
- s. 44: Obrázek by měl být „samostatný“ – vysvětlení symbolů S, C a O, které je uvedeno v textu, by mělo být uvedeno i v popisu obrázku.
- s. 46: Za příkazem CoDaDendrogram není =.
- s. 47: Popis CoDa-dendrogramu není snadný a čtenář neznalý tohoto nástroje zůstane v nejistotě, zda správně pochopil, co text říká.
- s. 51: Zakroužkování v dendrogramu je nanejvýš zavádějící. Týká se to zejména úplně nesmyslně vytvořeného červeného kroužku. Německo (v kroužku) má blíže ke státům mimo kroužek (Belgie-VB) než k ostatním z kroužku. Podobně např. Finsko se v procesu shlukování přiřadí dříve ke skupině Lucembursko-VB než k Dánsku a Švédsku. Podobně za „tradiční“ bychom místo modře označených států mohli stejně dobře označit i jiné shluky, např. Řecko, Itálie, Slovinsko, Maďarsko a Rakousko. Vertikální osa zde nemá význam.

V Olomouci 12. 5. 2017
Ondřej Vencálek