



POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Jméno studenta: Bc. Pavel Vojtas

Název práce: Vybrané metody analýzy vícerozměrných dat v MATLABu

Autor posudku: prof. RNDr. Hana Skalská, CSc.

Cíl práce: Analýza vícerozměrných dat se zaměřením na vybrané analytické metody a jejich využití pro různé typy úloh z různých aplikačních oblastí. Seskupení vybraných metod do logických celků pomocí grafického uživatelského rozhraní a tím jejich využití zefektivnit.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)			
	A	C	E	F
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Dílicí připomínky a náměty:

Cíl práce v kapitole 2 je pouze obecný, rozpor anotace s výsledkem práce – výsledky nejsou porovnány se specializovaným SW. V kapitole 3 postrádám upřesnění (přehlednější strukturovanou formou) typů úloh a metod, na které se práce konkrétně zaměří, místo obecného výčtu metod a konstatování u určitých typů, že se jimi práce zabývá.

Teoretický základ vybraných metod v kapitole 4 mohl být doplněn o přehled řešených úloh, ve kterých budou dané metody v práci využité. Vhodným doplňkem popisovaných metod by byl preciznější popis. Vazba metody na jejich konkrétní aplikace v práci by přispěly k přehlednosti.

Kapitola 5 popisuje reprezentaci výpočtů v prostředí Matlab, aplikuje výpočty vícerozměrných deskriptivních statistik a vizualizace na vícerozměrná data. V příkladu mění značení, zavedené v kapitole 4 (str. 9, 11 a 12). V kapitole 4 značí m počet pozorování, v kapitole 5 (str. 24) m označuje dimenzi vektoru pozorování (vhodnější by bylo obvyklejší p nebo k), n je počet pozorování. Potom ve vzorcích 5.6, 5.11 a 5.12 (str. 25) nastává pochybnost o správnosti zápisu.

Kapitola 6 popisuje vlastní aplikace. V příkladu 6.1.1. jsou data zastoupená absolutními četnostmi (počty cizinců, trestných činů, sňatků), která by bylo vhodné převést před analýzou na relativní ukazatele (vzhledem k celkovému počtu obyvatel vymezeného věku), aby hodnoty byly navzájem srovnatelné mezi okresy. Význam grafů 6.4 a 6.5 (str. 36 a 37) mohl být lépe vysvětlený (z grafů je vidět, že hlavní komponenty jsou nekorelované, což je vlastnost metody). Pro příklad 6.3.1 také platí poznámka (vhodnější relativní ukazatele) k příkladu 6.1.1, zde by volba *mohla* záležet na cíli shlukování.

Ve výsledku příkladu 6.3.2 (klasifikace dokumentů) na straně 52 není jednoznačná klasifikace do dvou shluků. Obrázek 6.12 (str. 52) ukazuje velký rozdíl na druhé komponentě mezi objekty PCAF a PCAV (tedy tři kategorie), tři shluky by vznikly také volbou 0,2 pro minimální vzdálenost (obrázek 6.11).

Připomínky formální a gramatické povahy.

Poměrně častými prohřešky jsou nevhodně zvolená interpunkce, ponechaná osamocená hláska na konci řádků, nevhodné slovní tvary nebo nevhodná spojení. Například: *třech* metod místo *tří* metod (str. 3), *hodnoty dat spolu provázané* (str. 4), neúplný nadpis na straně 5 (chybí *analýzy*), v celé práci *předúprava dat* (vhodnější *předpříprava*), ... *tyto slova ...*, *množině se ...říká* (str. 7), pojem *QR* není upřesněný (str. 8), předpoklady v definici 4.4 nejsou úplné, ... *to je vlastnost...* (str. 10) vhodnější například *což je vlastnost, ... po prvních ..* (? čem) str. 14 dole, dokument ...*se rozseká* na ... (str. 50), *MAT-LABovských* (str. 53), názvy metod by měly začínat malým písmenem (str. 55).

Místo tvaru *spočítat* (v celé práci) je možná vhodnější tvar *vypočítat* nebo *počítat* (celá kap. 5), dále není jasný význam pojmenování *dvojný graf* (např. 6.4, 6.5).

Umístění literatury by bylo vhodnější až na konci práce, za přílohami.

Celkové posouzení práce a zdůvodnění výsledné známky:

Práce se zabývá využitím širokého spektra úloh vícerozměrné deskriptivní statistiky, singulárního rozkladu, metody hlavních komponent, faktorové analýzy a shlukové analýzy. Kromě základní teorie popisuje a dokumentuje vlastní aplikaci, vytvořenou v prostředí Matlab.

Přínosem práce je sestavený přehled pokročilých metod s aplikacemi a stěžejní částí je vývoj objektově orientovaných komponent a grafického rozhraní pro Matlab.

Celkovou náročnost práce ukazuje také 24 citovaných a využitých odborných zdrojů.

Přílohy obsahují uživatelské příručky pro vytvořené aplikace a jsou v komprimovaném tvaru doloženy jako součást práce. Funkčnost aplikací zcela odpovídá popsaným úlohám.

Práce se zabývá problematikou a tématy nad rámcem obsahu studovaných kurzů.

Práci jako celek hodnotím výborně. Část bodů strhávám za drobnější problémy a nepřesnosti.

Otázka k obhajobě:

Které faktory nebo okolnosti rozhodovaly o volbě komponent (parametrů analýz), implementovaných do grafického rozhraní?

Práci doporučuji k obhajobě.

Navržená výsledná známka: B - výborně-velmi dobře

V Hradci Králové, dne 23. května 2017

podpis