

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Analýza metod a postupu pro zpracování dat z dopravních průzkumů
Jméno autora:	Martin Šír
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta dopravní (FD)
Katedra/ústav:	Katedra aplikované matematiky
Oponent práce:	Vladimír Socha
Pracoviště oponenta práce:	Katedra letecké dopravy

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadanie práce je formulované štandardne, t.j. kladie požiadavky na vytvorenie rešerše, výber a zdôvodnenie využitia vhodných štatistických metód pre spracovanie dopravných prieskumov, a aplikáciu týchto metód na reálnych dátach. Zadanie je formulované tak, aby poskytlo dostatočný priestor pre uchopenie tejto témy čím kladie požiadavky na kreativitu študenta. Je náročné ohodnotiť náročnosť zadania, to je možné považovať aj ako „jednoduché“ tak „mimoriadne náročné“, a to v závislosti na tom, ako je téma uchopená.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadanie je možné považovať za splnené.	

Zvolený postup řešení	částečně vhodný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Študent aplikoval základné štatistické metódy na dáta dotazníkového prieskumu. Očakával by som ale viac, viz. nižšie.	

Odborná úroveň	E - dostatečně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Základným problémom predloženej práce je pomerne skromná vlastná práca autora. Autor predviedol praktickú aplikáciu štatistických metód na reálnych dátach na 10 stranách v kapitole 5. Kapitoly 2,3,4 predstavujú zoznam štatistických metód a prístupov, ktoré podľa môjho názoru nemajú v kontexte práce žiadnu informačnú hodnotu. Jedná sa o všeobecne známe, jednoducho dohľadateľné a v praxi využívané základné metódy hodnotenia dát. Kapitola 3 je prepisom 5tich zdrojov, pričom doslova kopíruje osnovu prednášok Dr. Pecherkové a dokonca využíva aj grafiku z týchto prezentácií. Obdobne je tomu aj v Kap. 4.	
Vo svojej podstate sa tieto kapitoly v práci mohli nahradiť jednou vetou a referenciou, napr.	
„Pre prehľad štatistických metód vid'. Publikáciu „Přehled statistických metod“ [XX]“	
[XX] HENDL, Jan. <i>Přehled statistických metod: analýza a metaanalýza dat.</i> 3., přeprac. vyd. Praha: Portál, 2009. ISBN 9788073674823.	
Od prvých kapitol som očakával konkrétny a možno aj kritický prehľad využívaných metód pre spracovanie dotazníkov v doprave. Diskutovanie publikovaných prác, aké dotazníky sa používajú a ako sú štruktúrované, aký typ dát je najčastejšie spracovávaný, čo sa ostatní autori snažia analýzou adresovať. Takto by autor možno zistil, že je možné aplikovať metódy ako Korešpondenčná analýza, hierarchický clustering, Regresné stormy, atp...	

Kapitola 6. v práci nedáva žiaden zmysel.

Vlastná aplikácia štatistických metód na dotazníkové šetrenie prezentovaná v kapitole 5 je len akousi ukážkou. Autor si z celkového dotazníku vybral jednu otázku, i.e. do akej miery respondentov stresuje cestovanie verejnou dopravou a ukázal postup testovania nezávislosti cez Chi-kvadrát a využitie Wilcoxonovho testu.

V ďalšej časti je prezentovaná, ukážka Kruskal-Wallisovho testu. Tu autor tvrdí, že nemôže ukázať dáta, a to vzhľadom na ich veľkosť. Pritom sa ale tieto dáta zhrnúť v tabuľke (jedná sa o početnosti odpovedí ano, nie). Každopádne nemyslím si, že aplikácia K-W testu na tieto dáta je možná. Autor tu pravdepodobne porovnával 4 singulárne hodnoty medzi sebou (viz otázka oponenta).

Kapitola 5.1.6 neadresuje absolútne nič. Nerozumiem prečo v práci je.

Autor taktiež aplikoval PCA analýzu na jednu z multikriteriálnych otázok. Tu som nepochopil s akými dátami študent vlastne pracuje (viz. otázka oponenta). Pravdepodobne sa jedná o ohodnotenie jednotlivých kategórií na definovanej škále a to 445 respondentmi. U PCA analýzy mi chýba Score plot a loading plot - aspoň pre prvé dve hlavné komponenty (viz. otázka oponenta).

Následne autor prezentuje ukážku logistickej regresie a multinomického regresného modelu.

Celé diskutované spracovanie je realizované v prostredí matlab. Tu vyvstáva otázka, do akej miery bolo uvedené zložité vykonať, keďže Matlab statistical toolbox má všetky tieto metódy implementované aj s generáciou výsledkov, a teda by sa mohlo jednať o cca 10 riadkov kódu (zahŕňajúc načítanie a možno mierne predspracovanie dát).

Odborná stránka práce je teda postavená na jednoduchej ukážke aplikovania elementárnych štatistických metód na reálne dotazníkové dáta. Je škoda, že Autor neaplikoval aj iné metódy, a nevytvoril prehľad metód a vhodnosť ich aplikácie na konkrétne typy dát.

Formální a jazyková úroveň, rozsah práce

E - dostatečně

Posuďte správnosť používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Formálna úroveň práce je podľa môjho názoru na hranici prijateľnosti. Práca nemá štandardné členie IMRaD berúc do úvahy princípy akademického písania.

Práca je zbytočne členitá, kde niektoré kapitoly/podkapitoly/podpodkapitoly/podpodpodkapitoly... ani neobsahujú dostatok textácie. Niekde autor vytvára podkapitoly o jednej vete, inde využíva „bullet listy“. Uvedené potom pôsobia dojmom nedostatočnej jazykovej vybavenosti autora.

Kapitola 2 obsahuje 2 vety...

Obrázky a tabuľky nie sú referované v texte, a teda čitateľ ani nevie kedy na nich má upriamiť svoju pozornosť. Majorita grafiky je prevzatá, aj napriek tomu, že autor mohol využiť vlastné príklady histogramov, boxplotov, spojitých grafov atp., keď už pre celé vypracovanie práce používa Matlab. Umiestnenie obrázku 2 alebo tab. 1 nepôsobí dobre (odsadenie). Niektoré grafy (schémy) mohli byť zlúčené do jedného obrázku a označené ako A,B,C...

Tabuľka 3, 4 a 7 je totožná. Tabuľka 5 a 6 je totožná.

Rovnice nie sú číslované a u majority chýba popis premenných.

Práca v podstate neobsahuje žiadnu relevantnú rešerš, resp. analýzu súčasného stavu a ani diskusiu.

Autor pri písaní používa pasív, prvú osobu jednotného čísla ako aj prvú osobu množného čísla.

Výběr zdrojů, korektnost citací

F - nedostatečně

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

V práci je použitých 29 zdrojov, z čoho je možné považovať za relevantný zdroj iba zdroj [4]. Všetky ostatné zdroje predstavujú Wiki, prednášky a web. Nespochybnujem správnosť týchto zdrojov, avšak podľa môjho názoru by VŠKP mala pracovať s relevantnými zdrojmi, i.e. články, knihy, resp. so zdrojmi ktoré prešli nejakou formou kontroly, napr. recenzným konaním.

Nezdá sa mi úplne správne, aby VŠKP bola prepisom informácií umiestnených na Wikipédii, prednášok Dr. Pecherkové a webu.

Citovanie v texte je vykonané formou, kedy autor umiestňuje všetky citácie za odstavce. Pri bližšom pohľade je potom zjavné, že majorita textu práce je preparafrázovaný a ozdrojovaný cudzí text. Takýto spôsob citovania nie je správny.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Práca predstavuje kompendium štatistických metód s ukázkou aplikácie na dátach dotazníkového šetrenia.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

- 1) Vysvetlite aplikáciu KW-testu popísanu v kapitole 5.1.5.1.
- 2) Ukážte vstupné dáta využité pre PCA. Ukážte a popíšte Score a loading plot (môže byť aj formou bi-plot)

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **E - dostatečně**.

Datum: 27.8.2024

Podpis:

