

Posudek oponenta diplomové práce

Autor/autorka práce: David Peterka

Název práce: Statistické metody využitelné při analýze medicínských dat

Obsah práce

Student ve své práci využívá data zdravotnické záchranné služby Plzeňského kraje týkající se výjezdů k pacientům, kteří prodělali cévní mozkovou příhodu. Jako stěžejní část práce považuji kapitoly 4 a 5 poskytující přehled běžných statistických testů a kapitolu 6 se statistickou analýzou dat. Vzhledem k mohutnosti kapitoly 6.3 Testování dat (stránky 26-85) považuji za nešťastné, že interpretace výsledků není souhrnně v samostatné kapitole.

Text práce odpovídá 60 normostranám.

Kvalita řešení a dosažených výsledků

V práci je předložen ucelený statistický pohled na problematiku se zaměřením na hypotézy ohledně časů dojezdů, ošetření a transportu. To vyplývá z požadavků záchranné služby a Iktového centra v Plzni, jelikož doba, za kterou se pacienti s mozkovou příhodou dostanou do nemocnice, je kritickým faktorem úspěšnosti léčby. V práci jsou srovnávány data na úrovni okresů a na úrovni jednotlivých výjezdových základů v daném okrese.

Práce vychází ze tří souborů s tabulkovými daty za roky 2016, 2017 a 2018. V textu není vysvětlen význam jednotlivých atributů primárních dat. Student v rámci předzpracování soubory očistil od záznamů s chybějícími údaji a zařadil je do jím vytvořené databáze v MS Access. V databázi vytvořil množinu výběrových dotazů k získání podmnožin dat pro jednotlivé statistické testy. V práci postrádám zmínku o dalších kontrolách kvality dat. Není zřejmé, zda např. student ověřoval, že se v datech nevyskytují fakticky nemožné údaje, jako např. časový nesoulad v posloupnosti činností. Nezodpovězenou otázkou dále je, zda bylo bezpodmínečně nutné odstranit data s některou chybějící hodnotou. Dokáží si představit, že výjezd, u kterého chybí pouze čas předání pacienta do nemocnice, může být použit pro statistiky týkající se dojezdové doby k pacientovi. V práci není uveden přehled, jaké atributy nejčastěji chyběly, ani není diskutováno, co tento stav mohlo způsobit.

Funkcionalitu vytvořené databáze nemohu ověřit, jelikož vytvořená makra končí run-time chybami. Nejsem schopen posoudit, zda je problém způsoben nestandardní konfigurací počítače nebo špatnou implementací. Jednotlivé formuláře a dotazy jsou ale funkční.

Formální úroveň

Množství překlepů je vzhledem k rozsahu práce zanedbatelné. Reference nejsou číslovány podle prvního výskytu v textu, nýbrž podle abecedního pořadí jmen autorů. Nestandardní je i umístění seznamu zkratk za úvodní kapitolou. K dosažení lepší čitelnosti textu by bylo vhodné, aby autor explicitně zmínil testované hypotézy na samostatném řádku a neuváděl je pouze ve volném textu.

Pozitivně oceňuji uvedení seznamů obrázků a tabulek, které usnadňují orientaci v textu.



Práce s literaturou

Student pracuje s 28 zdroji, které jsou relevantní k řešenému problému. Většina zdrojů je použita k vybudování teoretického základu ke statistickým testům. U některých zdrojů by bylo vhodnější použít primární zdroje, např. pro texty zákonů (značky [1], [27]) citovat přímo ze Sbírký zákonů nebo pro demografická data (značka [15]) využít publikace Českého statistického úřadu.

Splnění zadání

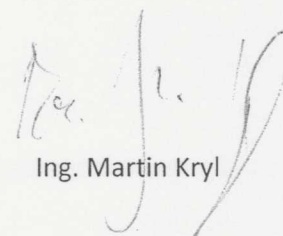
Zadání bylo splněno s menšími výhradami. Součástí práce bylo mimo jiné uložení dat ve vhodném formátu a jejich očištění. Z textu je obtížné dohledat, jaké činnosti proběhly v souvislosti s čištěním dat a jaký měly efekt. Dále zhodnocení možností využití získaných možností je velmi stručné.

Dotazy k práci

Jaké výhody má ve vašem případě použití MS Access databáze k uložení a předzpracování dat oproti provedení těchto činností v prostředí MS Excel?


Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 3. 6. 2019



Ing. Martin Kryl

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky
①


SOUHLASÍ
S ORIGINÁLEM