

Posudek oponenta diplomové práce

Autorka práce: **Jana Koukalová**

Název práce: **Statistické zpracování dat ruční antropometrie pro různé populace**

Hlavním cílem DP bylo vytvořit nástroj pro statistické zpracování dat pro účely studie zabývající se antropometrickými údaji o lidských rukách v české populaci. S využitím R Shiny studentka vytvořila webovou aplikaci určenou pro statistické zpracování dat z podobných studií.

Obsah práce

- Struktura DP je přehledná, jednotlivé kapitoly na sebe logicky navazují.
- Po obsahové stránce text často obsahuje zbytečně dlouhé textové pasáže bez přidané hodnoty, na druhou stranu detailnější popisy důležitých částí často chybí (sekce 3.3, 4.2, 5.3).

Kvalita řešení a dosažených výsledků

- Kompilační část Metodologie (sekce 3) je popsána většinou pouze slovně, s minimem matematických a statistických vzorců, zejména sekce 3.3 by si zasloužila detailnější rozpracování jednotlivých testů a metod.
- V části popisující vlastní programové řešení (sekce 4) studentka zvolila (veskci 4.2) dle mého názoru trochu nešťastně pouze popis I/O operací (načtení dat z Excelovských souborů, příprava výstupních grafů) a operací pro práci s uživatelským rozhraním (`observeEvent`) apod. částí z `main.R`, namísto aby poreferovala o implementaci jednotlivých statistických metod (např. z `pokrocilaAnalyza.R`).
- V části popisující vlastní statistickou analýzu případové studie (sekce 5) chybí (v seckci 5.3) konkrétní ukázka vstupního datového souboru před a po pročištění. Jaké hodnoty obsahovalo těch 42 záznamů odstraněných metodou MAD (resp. těch 29 záznamů u 3-sigma)? Zbylé části popisující výsledky jsou až na nečitelnost obrázků a tabulek (viz níže) provedeny velmi pěkně.

Formální úroveň

- Práce obsahuje poměrně velké množství překlepů, nelogických a nepřesných formulací, namátkou např.
 - o str. 9, ř. 9: „Pro svou názornost a intuitivnost je dospěno výsledku snadno.“
 - o str. 11: nepřekná sazba matematických vzorců v TeXu,
 - o str. 36: chybný popis tabulky 5.3, apod.
- V textu jsou často používané zkratky bez předchozího zavedení (např. v sekci 3.3.4), jejich nesetříděný seznam je uveden až v dodatku A.
- Obrázky 5.2-5.6 a 5.8-5.19 jsou naprosto nečitelné, celkem jde o 16 obrázků z 24, zobrazovat by měly zejména vlastní statistické výsledky.
- V celém textu je nekorektně vysázené české písmeno ý s obráceným apostrofem.
- Nevhodný font je použit i pro sazbu zdrojových kódů (není monospace), resp. velikost písma u tabulek, nečitelné jsou tabulky B.1.-B.29, celkem jde o 29 tabulek z 31, zobrazovat by měly zejména vlastní statistické výsledky.

Práce s literaturou

- DP obsahuje celkem 45 většinou relevantních referencí, které jsou v textu korektně citovány.
- Je trochu zarážející, že studentka po absolvování celé řady statistických předmětů zvolila jako referenci pro základní popisné charakteristiky (sekce 3.2.1) knihu [21] *Statistika bez předchozích znalostí*.

Splnění zadání

- Splněno bez výhrad.

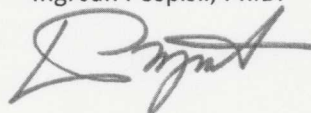
Dotazy k práci

- Detailně popište použití PCA (vč. ověření všech předpokladů) na konkrétních datech z Vaší antropometrické studie. Popis doplňte o názornou ukázkou výpočtu v části Advanced Statistics Vašeho programu.
- Detailně popište, jak jste postupovala při analýze rozptylu u dat s šířkou a obvodem dlaně, které nejsou normálně rozdělené (viz. např. obr. 5.3).
- Popište, co je p-hodnota testů a co nám na konkrétních příkladech z Vaší studie říká.

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

V Plzni 31.5.2019

Ing. Jan Pospíšil, Ph.D.



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



**SOUHLASÍ
S ORIGINÁLEM**

①