

POSUDEK OPONENTA NA DIPLOMOVOU PRÁCI

BLANKA SUDOVA: ANALÝZA PŘÍJMOVÝCH NEROVNOSTÍ

Splnění cílů práce

- nadstandardně
- velmi dobře
- splněny
- s výhradami
- nebyly splněny

Odborný přínos práce

- nové výsledky
- netradiční postupy
- zpracování výsledků z různých zdrojů
- shrnutí výsledků z různých zdrojů
- bez přínosu

Matematická (odborná) úroveň

- vynikající
- velmi dobrá
- průměrná
- podprůměrná
- nevyhovující

Věcné chyby

- téměř žádné
- vzhledem k rozsahu přiměřený počet
- méně podstatné, větší množství
- podstatnější, větší množství
- závažné

Grafická, jazyková a formální úroveň

- vynikající
- velmi dobrá
- průměrná
- podprůměrná
- nevyhovující

Autorka z dat o příjmech jednotlivých zaměstnanců vybraných firem v Argentině vypočítává indexy nerovnosti (Giniho, Theilův) v příjmech celého vzorku, resp. v jednotlivých podmnožinách (podle pohlaví, regionu, oboru, starší/mladší, ve firmě delší/kratší dobu). Zároveň kvantifikuje nerovnosti mezi podmnožinami, jako měřítko zde zavádí plochu pod grafem zobecněných relativních příjmových rozdílů, resp. plochu pod pravděpodobnostním grafem. Kromě toho logistickou regresí odhaduje podíl žen mezi zaměstnanci v závislosti na příjmu a dalších uvedených veličinách.

Práce je mi sympatická, nicméně musím uvést několik bodů, které jinak dobrý dojem kazí.

1. Práce vychází z dat, která *nejsou součástí* práce, není tedy jak ověřit správnost výsledků. Čekal bych, že v textu bude o to podrobnější popis, např. uvedena alespoň ukázka zdrojových dat. Lepší představě, a to i z hlediska rozebíraného tématu, by pomohlo např. vykreslení histogramů příjmů v jednotlivých zkoumaných skupinách (jediným histogramem rozdělení příjmů je celkový histogram před odstraněním dat), systematické uvádění charakteristik jednotlivých skupin (uvedeny jsou vlastně jen v úvodu u žen a mužů, pak už asi jen průměry u oborů).
2. Při zavádění indexů *PGI* a *PRI* pro stochasticky neporovnatelné populace *není položena* otázka, zda je index dobře definován — zda po prohození skupin skutečně vyjde kladný, popř. co když kladný vyjde v obou případech.

3. Otázkou je *reprezentativnost* zkoumaného výběru a tudíž aplikovatelnost vyslovených závěrů. Uvádí se pouze to, že jde o zaměstnance soukromých firem. Přitom z občasných údajů uvedených v textu (výběr firem, počty obyvatel regionů vs. počty ve výběru, počty dle oboru) takové podezření vzniká, při výpočtu zohledněno není. Navíc vyřazeny byly údaje o příjmech více než poloviny zaměstnanců ve výběru (jsou rozdělení příjmů a míry nerovnosti u vyřazených stejně jako u nevyřazených?).
4. Uváděné hodnoty indexů jsou vlastně bodovými odhady neznámé teoretické hodnoty, měla by se tudíž vzít v úvahu i chyba odhadu. Rozumím, že její odvození by nejspíš nebylo jednoduché (čili neuvedení např. intervalů spolehlivosti práci nevytýkám), ale ne vždy je *neznámé spolehlivosti* odhadu přizpůsobena formulace závěrů.
5. Přepočítání pravděpodobnosti ženy ze zaměstnanců na celou populaci je podivný. Pokud osoba nemá příjem, nelze příslušnou pravděpodobnost určit. Přepočítání neodpovídá ani případnému hypotetickému předpokladu, že rozdělení příjmů všech žen by bylo stejné jako u těch ve výběru.

Práce je psána přiměřeně, i když na některé drobnější chyby v prezentaci narazíme. Například: Ve výkladu při zavádění indexů míchá teoretické a výběrové definice (v základním a výběrovém souboru), u percentilů přechází mezi $x \in \langle 0, 1 \rangle$ a $x \in \langle 0, 100 \rangle$ a nespecifikuje, jakou definici kvantilu používá. Jak je definováno TI, když některé y_i bude nulové (nulový příjem je v příkladech explicitně uváděn)? Křivkou v obrázcích u PRI není graf PR_x , ale graf funkce k ní „inverzní“, a i popis na konci str. 17 je matoucí. Obr. 1.1 ukazuje záporné příjmy? $F(y)$ bude rostoucí (str. 7) jen na $y > 0$, v (2.9) na str. 9 je nevhodné dvojí i . Vodorovné osa v grafech PGI je popsána jako Percentil x , je to ale přímo x . Proměnné věk a doba (str. 44) jsou multinomické? Mezi sporné bych zařadil označení výběru za náhodný (str. 2), zvolenou implementaci výpočtu PGI a PRI (volil bych jinak než zrovna percentily). Opatrnější formulace by byly vhodné při interpretaci výsledků, pro pojmenování podle délky (zatímního!) působení ve firmě bych hledal lepší označení než loajalita.

Hodnotím práci známkou *velmi dobře* a doporučuji k obhajobě.

Otázky k rozpravě při obhajobě:

- Je při „chybné“ úvodní volbě „horší skupiny“ nutné přepočítávat index PRI, resp. index PGI? Souvisí spolu nějak hodnoty indexů spočtené pro každou skupinu zvlášť a hodnota indexu porovnávajícího obě skupiny? Nulovost indexu PRI (PGI) a shoda rozdělení v porovnávaných skupinách?
- V práci se popisuje aproximace křivek pomocí hodnot zjištěných pro „celočíslné“ percentily. Jakou definici těchto celočíselných percentilů používáte a proč jsou použity právě celočíselné percentily?
- Šel by k vyjádření nerovnosti příjmů v populaci použít např. koeficient variace (směrodatná odchylka dělená průměrem)? Dopadlo by porovnání nerovnosti mezi skupinami stejně, kdyby se porovnály pouze průměry?



MICHAL FRIESL

Plzeň, 6. června 2016.