

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Zobecněný Wilcoxonův test pro cenzorovaná data

Autor: Michaela Vařejková

SHRNUTÍ OBSAHU PRÁCE

Práce je členěna do tří kapitol. V první kapitole je představen jedno i dvouvýběrový Wilcoxonův test, jsou shrnuty jejich základní vlastnosti. Dále je zde diskutována problematika cenzorování. Hlavní částí práce je druhá kapitola, která zavádí zobecnění dvouvýběrového Wilcoxonova pro použití s cenzorovanými daty. Opět jsou uvedeny základní vlastnosti nutné pro použití testu s reálnými daty. Ve třetí kapitole jsou uvedeny výsledky simulační studie mající za cíl empiricky prozkoumat sílu, resp. dodržování hladiny zobecněného i nezobecněného Wilcoxonova testu v situaci, kdy je tento použit „naivním“ způsobem při ignorování cenzorovaných pozorování.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Téma umožnilo přiměřeným způsobem rozvinout poznatky získané v povinných kurzech, zejména pak v přednášce *Matematická statistika 1*. Hlavním cílem práce bylo odvodit, za pomoci doplňkové literatury, statistické vlastnosti zobecněného Wilcoxonova testu. Tohoto cíle bylo dle mého názoru dosaženo, byť se značnými výhradami, viz dále.

Vlastní příspěvek. Vlastní příspěvek autorky spatřuji v poskytnutí poměrně přehledného shrnutí kroků nutných k odvození základních vlastností testových statistik Wilcoxonova testu s cenzorovanými daty. K tomuto mám nicméně poměrně značné výhrady, viz matematická úroveň práce a seznam připomínek. Plně vlastním příspěvkem autorky je potom provedení simulační studie a prezentace jejích výsledků ve třetí kapitole.

Matematická úroveň. Matematické části textu jsou uváděny vesměs ve formě obvyklé pro matematický text. Nicméně, jak uvádím níže, hlavní část práce, kterou je druhá kapitola, sleduje poměrně těsně primární Gehanovy zdroje a jakmile se situace jenom mírně zesložití, mám obavu, že autorka značně sklouzává k překladu anglického textu, kterému po matematické stránce ne zcela rozumí. V principu jediná část, která v nějaké formě jde nad rámec zmíněných zdrojů, je (relativně jednoduchý) důkaz Věty 2 (jejíž tvrzení je v článku uvedeno jako „*It is easy to show that...*“)

Práce se zdroji. Zdroje, z nichž práce vychází, jsou řádně a též formálně správným způsobem citovány. Nutno však konstatovat, že hlavní část práce, 2. kapitola, sleduje svoji strukturou poměrně těsně řádně citované primární zdroje, kterými jsou anglicky psané Gehanovy články z roku 1965.

Formální úprava. Formální úprava práce je na dobré úrovni. Jako výtku bych však uvedl fakt, že tabulky i obrázky jsou uváděny uvnitř textu, což do jisté míry znepřehledňuje jeho čitelnost. Výskyt pravopisných chyb či překlepů je nízký, rozhodně v míře obvyklé pro tento druh práce.

1. Hlavním výstupem oddílu 2.1.1 by měl být rozptyl testové statistiky za nulové hypotézy. Na základě Gehanova článku autorka navrhuje nejprve podmiňovat „schématem P“. Na vlastním hezkém příkladě ilustruje, jak schéma P vznikne v konkrétní situaci. Nicméně hlavní výstup, rozptyl testové statistiky, je uveden jako pouhý opis výrazu z původního článku (kde lze odvození nalézt v appendixu). Z jakého důvodu se autorka ani nepokusila o přepis tohoto odvození (s doplněním méně zřejmých kroků)? Můj hlavní dotaz k této části: Jakým způsobem lze odvozený rozptyl podmíněný schématem P použít k výpočtu rozptylu nepodmíněného, který nás zejména zajímá, nepletu-li se.
2. Definice 2 na str. 13 je zřejmě motivována textem appendixu B Gehanova článku. Ze značení použitého autorkou je nicméně poměrně obtížné poznat, kdy se jedná o vektory a vlastně není ani zřejmé, jestli se vůbec jedná o vektory. Veličiny x a y v argumentech funkcí F_{X_α} a F_{Y_β} mají být x_α a y_β ? Distribuční funkce F_{X_α} odpovídá jedno nebo dvourozměrnému rozdělení?
3. Str. 13: Argumentace „*Víme, že pokud pro $n \rightarrow \infty$ existuje $\lim n/m \dots$* “ mi není zřejmá. Plyne toto z nějaké věty či tvrzení uvedených v práci Lehmann (1951), na kterou je polopaticky odkázáno? Hned dále jsem zcela zmaten vyjádřením „ $x_\alpha = (x_\alpha^{(1)}, x_\alpha^{(2)})$, kde $x_\alpha^{(1)} = x_i, x'_i$.“ Konečně mi není vůbec jasné, proč je statistika U stejná jako $W/(nm)$ a proč jsou momenty zde uvedené „dobře definovány“, resp. konečné. Mám obavu, že tato část bakalářské práce až příliš těsně sleduje Gehanovu argumentaci bez jakékoliv snahy o porozumění.
4. Oddíl 2.2 je prakticky celý výtahem (přeloženým do češtiny) z druhého Gehanova článku z roku 1965.

ZÁVĚR

Zejména s ohledem na obsah druhé, stěžejní kapitoly, který jde jenom velice minimalisticky nad rámec výchozích zdrojů a co horšího, navozuje dojem bezmyšlenkovitého opisu stěžejních částí, pouze se značnými výhradami a velice hraničním způsobem **doporučuji** práci uznat za bakalářskou na oboru *Obecná matematika*. K faktu práci ještě stále doporučit k uznání za bakalářskou mě potom vede třetí kapitola, jež je zpracována poměrně hezky.

V Kačlehách dne 22. srpna 2019

doc. RNDr. Arnošt Komárek, Ph.D.

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta Univerzity Karlovy