

Posudek oponenta na diplomovou práci
Vlastností dvoufázových testovacích procedur
Eliška Krausová

Diplomová práce se zabývá vlastnostmi dvoufázových testových procedur, tj. postupu, kdy se nejdříve testuje nějaký předpoklad a podle výsledku tohoto „předběžného“ testování předpokladu se vybere test hypotézy, která nás zajímá. Takovéto dvoufázové testování je velmi často doporučováno zejména v aplikované literatuře. Proto je pro statistickou praxi velmi důležité rozumět vlastnostem takových procedur.

Práce je slušně napsaná, i když má své formální nedostatky. Struktura práce je logická. Autorka prokázala, že porozuměla praktickým implikacím daného problému. V práci však postrádám solidní teoretické porozumění daným problémům.

Připomínky, návrhy, postřehy:

Kapitola 2 „Dvoufázové testovací procedury“.

Tato kapitola představuje některé dvoufázové procedury a formálně popisuje obecnou dvoufázovou proceduru.

- (1) Strana 8: Nemám možnost si to ověřit, ale předpokládám, že podmínka $n_1 \cdot n_2 > 10$ je asi překlep.
- (2) Strana 11: Obrázek 2.3 je velmi názorný. Myslím si však, že jej není nutné opakovat ještě na straně 19 a 23.
- (3) Strana 13: Jsou zde použity nezavedené symboly, což vadí zejména pro \bar{x} a také není v literatuře úplně jednoznačné používání symbolu $t_{n-1}(\alpha)$. Ostatní symboly jsou naštěstí standardní.
- (4) Strana 13: Zde se zřejmě jedná o jednovýběrový t -test a nikoliv o dvouvýběrový.
- (5) Strana 13: Asi by mělo být

$$\mathcal{F}_0^* = \left\{ F \in \mathcal{F} : P_F \left(\left| \frac{\bar{X} - \mu}{S} \sqrt{n} \right| > t_{n-1}(\alpha) \right) \leq \alpha \right\}$$

Kapitola 3 „Statistické testy“.

Tato kapitola popisuje některé základní jednovýběrové a dvouvýběrové testy. Bohužel, ani tato jednoduchá rešeršní kapitola se neobešla bez nepřesností a formálních nedostatků.

- (1) Strana 14: T-test nebo t -test?
- (2) Strana 14: Když už ne na straně 13, tak alespoň zde by měly být zavedeny symboly \bar{X} , S , t_{n-1} .
- (3) Strana 14: Pokud autorka tvrdí, že je známo, že t -test není příliš citlivý na předpoklad normality a nevysvětluje to blíže, tak by měl být uveden alespoň odkaz na literaturu. V opačném případě není ani dopátratelné, co je tou necitlivostí myšleno. Dodržování hladiny? Síla? A jaký typ odchylek od normality se myslí?
- (4) Strana 15: Po konstatování, že neparametrickou obdobou jednovýběrového t -testu je jednovýběrový Wilcoxonův test, následuje definice. Nikoliv však Wilcoxonova testu, jak by čtenář očekával, ale kritické hodnoty normovaného normálního rozdělení, která je však zapotřebí až o více než půl stránky níže.
- (5) Strana 16: Asi by mělo být $\bar{X} - \bar{Y} - (\mu_1 - \mu_2)$.
- (6) Strana 16: Slušelo by se zavést alespoň symboly S_X^2 a S_Y^2 .

- (7) Strana 16: Necitlivost dvouvýběrového t -testu na předpoklad normality - viz výše poznámka o jednovýběrovém t -testu.
- (8) Strana 17: Věta 3.6 je formulována poněkud vágně. Co to znamená, že „přibližně platí“?
- (9) Strana 18: Ve větě „Místo veličiny se zpravidla používá veličina“ asi vypadlo T_1 .
- (10) Strana 18: Není jasné, co se myslí testovou statistikou U . Ta nebyla definována.
- (11) Strana 19: Definici množiny rozdělení \mathcal{F}_0^* je zapotřebí opravit podobně jako v poznámce (5) ke Kapitole 2.
- (12) Strana 20: Asi není nutné popisovat Lillieforsův test normality, když tento test není v práci použit.
- (13) Strana 21 popis F -testu: Slušelo by se napsat, co dělat když $S_X^2 < S_Y^2$? Nicméně ani F -test není v práci použit, takže je otázka, proč ho vůbec uvádět.
- (14) Strana 22: Volba Moodyho testu se mně obzvláště v kontextu dvoufázových procedur nezdá vhodná. Tento test je zkonstruovaný pro detekci změny měřítka při stejném parametru polohy v obou výběrech. K zamítnutí však také může vést pouhý rozdíl v parametru polohy, při stejném měřítku pro oba výběry. Pokud k volbě tohoto testu vedla nějaká dvoufázová procedura uvedená v literatuře, je škoda, že to není v práci uvedeno.
- (15) Strana 22: Jsou myšleny přesné tabulky kritických hodnot?

Kapitola 4 „Asymptotické vlastnosti dvoufázových procedur“.

Tato kapitola patří k hlavnímu teoretickému přínosu práce. Autorka zde obecně odvodila asymptotickou sílu dvoufázového postupu. Škoda jen, že výsledků této kapitoly nebylo více využito při komentování výsledků simulací.

- (1) Strana 24: Chtělo by to být přesnější ohledně používání slova „konvergují“.
- (2) Strana 27: Asi by zde mělo být $\mathcal{G}_0^* \cap \mathcal{G}_0$
- (3) Strana 28: Nerozumím, proč by měla být asymptotická síla $\min(P_{zam}(G), P_{zam}(H))$. Já bych spíše řekl, že pro alternativu G je síla $P_{zam}(G)$ a pro alternativu H je $P_{zam}(H)$.

Kapitola 5 „Simulace“.

V této kapitole autorka pomocí simulací zkoumala hladinu a sílu některých dvoufázových procedur. I když se jedná o solidně odvedenou práci, chybí lepší teoretický náhled do výsledků.

- (1) Pokud rozdělení není symetrické (kolem nějakého bodu a), tak jednovýběrový Wilcoxonův není testem o mediánu X_1 , ale mediánu $\frac{X_1+X_2}{2}$. Dodržovat hladinu, ale stejně tento test nebude, protože $X_1 - \mu$ (ať už je μ cokoliv) nemá rozdělení symetrické kolem nuly.
- (2) Dvouvýběrový Wilcoxonův test není testem o mediánech rozdělení, ani testem shody rozdělení (jak se tvrdí např. na straně 47), ale testem o tom, zda $P(X_1 < Y_1) = \frac{1}{2}$. Shoda rozdělení za nulové hypotézy je přepoklad, aby pořadí byla za nulové hypotézy „exchangeable“ a mohli být stanoveny příslušné kritické hodnoty. Tento předpoklad by však šel odstranit.
- (3) Některé obrázky nejsou příliš povedené. To platí např. o Obrázku 5.5. Nevidím důvod, proč by obrázek nemohl být širší a ne tak vysoký. Širší by jistě také mohl být i Obrázek 5.4.
- (4) Strana 40, poslední řádek: Nerozumím přesně tomu, co myslí autorka slovy „její chování je optimální“.

- (5) Ani trochu se mně nelíbí procedura, která se na základě testu o shodnosti rozptylů rozhoduje mezi t -testem a Wilcoxonovým testem. Na straně 7 této práce je uvedeno, že tento postup je doporučen v knize Zvára (2004). V této knize je však pouze uvedeno: „*Dvouvýběrový Wilcoxonův (Mannův-Whitneyův) test je málo citlivý vůči nestejným rozptylům σ_x^2, σ_y^2 . Podobně jako dvouvýběrový t -test citlivě reaguje na posunutí (nestejně parametry μ_x, μ_y). Proto se používá místo dvouvýběrového t testu také tehdy, když jsou rozptyly v porovnávaných výběrech příliš nestejně.*“ Osobně bych se neodvažoval brát tuto poznámku jako doporučení. Je však docela možné, že doporučení tohoto typu se vyskytují i jinde v literatuře, protože panuje mýtus, že Wilcoxonův test „se dá použít vždy“, protože je přece „neparametrický“. Škoda jen, že žádná další reference v tomto smyslu již v práci není. Ve shodě s diagramy na stranách 9 a 10 by se spíše nabízelo na základě testu o shodnosti rozptylů volit mezi t -testem se stejnými či různými rozptyly.

Kapitola 6 „Závěr“.

Autorka zde shrnuje některá zjištění. Nicméně, pro čtenáře by bylo jistě velmi užitečné, kdyby blíže specifikovala kdy a kde zjistila „že dvoufázové procedury hladinu dodržují jen pokud...“ či „Na simulacích jsme ukázali, že předpoklad normality má smysl testovat jen pro malé výběry...“

LITERATURA

Zvára, K. (2004). *Biostatistika*. Karolinum, Praha.

Domnívám se však, že, i přes výše uvedené výhrady, předložená práce splňuje požadavky kladené na práci diplomovou, a proto ji doporučuji za ni uznat.

5. května 2010

Ing. Marek Omelka, Ph.D.
KPMS MFF UK