

## POSUDEK OPONENTKY BAKALÁŘSKÉ PRÁCE

**Název:** Párové testy a jejich interpretace

**Autor:** Haojun Zhan

### *Shrnutí obsahu práce*

Práce obsahuje přehled parametrických a neparametrických párových testů, které se používají pro testování hypotéz o středních hodnotách ve dvourozměrných náhodných výběrech, jejichž realizací jsou párová data. Parametrický párový t-test je porovnán s dvouvýběrovým t-testem, přičemž autor vysvětluje důsledky jejich záměny. U neparametrických testů je kladen důraz na různé způsoby jejich interpretace, což je věc, která může činit potíže při praktickém použití testů.

### *Celkové hodnocení práce*

**Téma práce.** Téma považuji za přiměřené pro bakalářskou práci.

**Vlastní příspěvek.** Základní teorie k popsaným testům je součástí českých učebních materiálů k výuce statistiky na bakalářském oboru Finanční matematika. Za vlastní příspěvek autora lze považovat podrobné pojednání o různých způsobech interpretace neparametrických testů a provedení jednoduché numerické studie v programu Mathematica.

**Matematická úroveň.** Text je psán přehledně a srozumitelně. Matematické pasáže (důkazy vět, odvozování v numerické ilustraci) jsou podrobně rozepsány. Odborná náročnost nepřekračuje úroveň běžné učební látky.

**Jazyková úroveň.** Práce je psána v českém jazyce, místy s neobratnými formulacemi a drobnými gramatickými chybami. Občasné tiskové chyby v doprovodném textu zásadně neovlivňují jeho kvalitu, tiskové chyby v matematickém textu zmiňuji níže.

**Formální úprava** je na uspokojivé úrovni, co se týče členění do kapitol, kvality textu i grafických výstupů.

**Práce se zdroji.** Zdroje ze seznamu literatury jsou v textu řádně citovány. Nepracovalo se také s textem Kulich: NMF301 Statistika pro finanční matematiky. Přehledový větník?

### **Připomínky.**

- 1) Str. 3, Úvod, 1. odstavec, též str. 5, 1. odstavec a dále: *statistický rozdíl* ... spíše by mělo být statisticky významný rozdíl.
- 2) Str. 3, Základní předpoklady, 2. odstavec: o náhodných veličinách je použit výraz *stejně nezávislé rozdělené* ... patrně nevhodný překlad independent identically distributed.
- 3) Str. 6-7: dvouvýběrový t-test je včetně testové statistiky uváděn za předpokladu obecně rozdílných rozptylů, ale pod názvem t-test se standardně uvádí varianta se shodným rozptylem ve výběrech a jinou testovou statistikou. Zde by byl vhodnější například název dvouvýběrový z-test (viz Kulich, M.: NMF301 Statistika pro finanční matematiky. Přehledový větník, KPMS MFF UK, 2014) nebo Welchův test (viz Zvára, K.: Základy statistiky v prostředí R, Karolinum, Praha 2013). Analogicky by jako párový t-test na str. 5 měl být označen model  $\mathcal{F}_t$  a model  $\mathcal{F}_z$  pak jako párový z-test.
- 4) Str. 13, poslední řádek: je-li použita substituce  $z = m_z + x$ , má pak být v integrálu  $F_Z(m_z + x)$ , následně v odečítaném integrálu v 2. řádku na str. 14 má být  $F_Z(m_z + z)$ .
- 5) Str. 14, úpravy pod (2.5), 2. řádek: nejasná znaménka u 3. a 4. sčítance.
- 6) Str. 14-15: v komentáři k obrázku 2.1 by mělo být vysvětleno použité rozdělení, které je naznačeno v popisech svislých os.
- 7) Str. 16: zatímco důkazy vět 1- 4 jsou podrobně rozepsány, patrně na základě studované literatury, důkaz věty 5, který je triviální a je součástí látky z Pravděpodobnosti pro finanční matematiky, je odkazem na nepříliš známou zahraniční literaturu.

- 8) Str. 18, kap. 3.1, 2. řádek odspodu: místo  $Z_i$  má být  $Z_i^*$ .
- 9) Str. 19, začátek kap. 3.2: píše se, že symetrie hustoty kolem nuly je předpokladem párového Wilcoxonova testu, což není v souladu s modelem  $\mathcal{F}$  na str. 18.
- 10) Str. 20, řádek 14-15: alternativa  $\delta_Z \neq \delta_0$  je chybně označena jako  $H_0$ .
- 11) Str. 22, Závěr, 2. odstavec: *test rovnosti středních hodnot dvou závislých pozorování ...* mělo by být test rovnosti středních hodnot obecně závislých složek ve dvourozměrném náhodném výběru.

### **Otázky.**

- 1) Str. 6: z čeho plyne konvergence v distribuci veličin  $T_n$ ? Je  $\alpha$  přesná hladina pro příslušný asymptotický test založený na  $T_n$ ?
- 2) Str. 10, řádek 6: píše se o symetrii hustoty binomického rozdělení ... jak je to myšleno?
- 3) Str. 10, řádek 12: Co znamená  $G_0(y_n - 1)$ ?

### **Závěr**

Předložená práce je přehledovým textem, který z valné většiny přebírá látku z české učebnicové literatury. Ačkoli posluchač splnil zadání práce, je škoda, že kromě teoretické části zde nebyla požadována či z autorovy iniciativy přidána část praktická, například ve formě aplikace jednotlivých testů na vhodná párová data nebo vyšetřování jejich empirické hladiny nebo síly. Na takovém úkolu by se student naučil něco nového a přinesl výraznější vlastní příspěvek. Vzhledem k nízké náročnosti a jazykovým i matematickým nepřesnostem hodnotím práci spíše jako podprůměrnou, nicméně doporučuji její přijetí k obhajobě.

RNDr. Jitka Zichová, Dr. KPMS MFF UK  
19.8.2024.