

## POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

**Název:** Pearsonův korelační koeficient a jeho využití ve statistice

**Autor:** Richard Németh

### SHRNUTÍ OBSAHU PRÁCE

Bakalářska práce pojednává o Pearsonovom korelačnom koeficiente, jeho empirickej verzii a základných štatistických vlastnostiach a možnostiach jeho využitia.

V úvode práce autor definuje Pearsonov korelačný koeficient, jeho teoretickú a empirickú verziu a uvádza niektoré základné tvrdenia. Následne odvodzuje niektoré štatistické vlastnosti, hlavne asymptotické rozdelenie empirickej verzie korelačného koeficientu, a to za rôznych predpokladov na rozdelenie náhodného výberu a tiež prezentuje niekoľko spôsobov ako tento výberový korelačný koeficient využiť k testovaniu nulovej hypotézy o nezávislosti dvoch náhodných veličín.

V záverečnej časti práce autor pomocou simulačnej štúdie porovnáva fungovanie rôznych testových štatistík v závislosti na rozsahu náhodného výberu a tiež rozdelenia, z ktorého náhodný výber pochádza.

### CELKOVÉ HODNOCENÍ PRÁCE

**Téma práce.** Náročnosť práce považujem za primeranú a zadanú tému vypracovanú v rámci predložených zásad pre vypracovanie.

**Vlastní příspěvek.** Práca má prevažne kompilačný charakter. Hlavný príspevok autora vidím v poslednej časti, kde pomocou simulácií porovnáva rôzne prístupy k testovaniu nezávislosti.

**Matematická úroveň.** Matematický text je písaný korektne, v rámci zaužívaných spôsobov a práca celkovo spĺňa predpoklady kladené na bakalársku prácu.

**Práce se zdroji.** Práca korektne cituje použité zdroje, neobjavil som v tomto smere žiadne nedostatky.

**Formální úprava.** Z formálnej stránky považujem prácu za vyhovujúcu a vypracovanú na veľmi kvalitnej úrovni.

### PŘIPOMÍNKY

1. Tvrdenie 1.6 na str. 2 je v uvedenom tvare nesprávne. Zároveň nesedí označenie a indexy.
2. na str. 3 sa asi myslí realizace  $t_n = T_n(\mathbf{X}, \mathbf{Y})$ , namiesto  $t_n = T_n(\omega)$ ; Zároveň nie je úplne jasné, aký kvantil má autor na mysli vo výraze  $|t_n| \geq u_{1-\alpha/2}$ ; vo všeobecnosti sa toto označenie používa pre kvantil štandardného normálneho rozdelenia  $N(0, 1)$ , v tomto prípade má ale štatistika  $t_n$  študentovo  $t$ -rozdelenie s  $n - 2$  stupňami voľnosti;
3. v sekcii so simuláciami trochu chýbajú vo výsledkoch smerodatne chyby u uvedených priemerných hodnotách testových štatistík; doplnilo by to celkový obraz o kvalite simulácií;

## NĚKTERÉ CHYBY A PŘEKLEPY

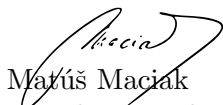
1. v práci som našiel len pár drobných preklepov, niektoré pôsobia naozaj zaujímavo (napr. "Pearsonův correlation coefficient" v anglickom názve, ...)
2. z práce nie je zrejmé, či  $(\sigma^*)^2$  a  $(\sigma^*(\rho))^2$  sú dve rôzne kvantity, alebo len dva rôzne označenia;
3. na str. 13 sa uvádza, že študentovo  $t$ -rozdelenie môžeme vďaka CLV pre veľké stupne voľnosti aproximovať normálnym  $\mathcal{N}(0, 1)$  rozdelením, to je ale skôr vlastnosť rozdelenia, než aplikácia CLV;
4. v použitom značení na str. 17 by možno bolo vhodnejšie vyjadriť závislosť veličín  $Q$  a  $U$  na  $n \in \mathbb{N}$ :  $Q_n \xrightarrow{H_0, d} \mathcal{N}(0, 1)$ , a analogicky aj pre  $U_n$ ;

## OTÁZKY

1. Ako by malo správne vyzerat' Tvrdenie 1.6 na strane 2?
2. V prípade simulácii sa v práci uvádza, že sa simuluje nahodný výber  $(X_1, Y_1)^\top, \dots, (X_n, Y_n)^\top$  z rozdelnia  $F$ . V niektorých prípadoch sa jedná o dvojrozmerné rozdelenie, v iných to je marginálne rozdelenie jednotlivých zložiek náhodného vektora  $(X, Y)^\top$ . Boli v simuláciach uvažované aj prípady, keď rozdelenie jednej a druhej zložky náhodného vektora nie je to isté?

## ZÁVĚR

Práci považuji za veľmi dobrou a doporučuji ji uznat jako bakalářskou práci.



Matúš Maciak

Katedra Pravděpodobnosti a Matematické Statistiky, MFF UK

10.06.2015