Posudek vedoucího bakalářské práce

Název: Statistické testy ve stratifikovaných kontingenčních tabulkách

Autor: Peter Vook

Shrnutí obsahu práce

Práce vychází z teorie dvourozměrných a zejména čtyřpolních kontingenčních tabulek, jež je přednášena v základním kurzu matematické statistiky. Zejména s ohledem na nutnost zavést potřebné značení je tato teorie stručně shrnuta v první kapitole předložené práce. Ve druhé kapitole autor zavádí klasické míry asociace používané v kontextu čtyřpolních tabulek a s využitím klasických vět odvozuje asymptotická rozdělení souvisejících odhadů. Stratifikované čtyřpolní tabulky jsou zavedeny v kapitole třetí, autor následně definuje související pojmy podmíněné nezávislosti a homogenních asociací. Ve čtvrté kapitole autor, opět s využitím obecných asymptotických vět odvozuje několik standardních testů používaných v kontextu stratifikovaných čtyřpolních tabulek. Jedná se zejména o testy podmíněné nezávislosti, (i) založený na asyptotickém chování logaritmických poměrů šancí, (ii) Cochranův-Mantelův-Haenszelův test založený na hypergeometrické reprezentaci. Stručněji se autor věnuje též Breslowovu-Dayovu testu homogenních asociací. Práce je zakončena menší simulační studií mající za cíl vyšetřit chování vybraných testů v případě konečných výběrů. Konečně je práce v apendixu doplněna o seznam nejdůležitějších používaných vět.

Celkové hodnocení práce

Téma práce. Téma bylo voleno tak, aby bylo možné přímé využití poznatků získaných v povinných kurzech, zejména pak v přednáškách *Matematická statistika 1 a 2*. Cílem bylo použít tyto poznatky a podrobně odvodit vybrané statistické testy. Dle mého názoru bylo tohoto cíle dosaženo.

Vlastní příspěvek. Kromě podrobného odvození testů podmíněné nezávislosti ve čtyřpolních kontingenčních tabulkách prezentuje autor výsledky simulační studie, jež má za cíl poodhalit chování vybraných testových statistik v případě konečných výběrů.

Matematická úroveň. Matematické části textu jsou uváděny vesměs ve formě definic a tvrzení při konzistentním používání zavedeného značení. Hrubé chyby, mohu-li soudit, se v práci nevyskytují. Autor se nicméně nevyhnul několika nepřesnostem, resp. opomenutím, jež mírně zhoršují dojem z práce.

- Na str. 4 a 5 je chybně zaměňován index k a n.
- Na str. 12 a 13 není zaveden význam symbolů n_{ijk} a μ_{ijk}.
- Na str. 14 se ne zcela korektně usuzuje z hodnoty bodového odhadu na podmíněnou nezávislost.
- Symbol $\overset{as.}{\sim}$ použitý např. na str. 15 není zaveden.
- Ve druhém a třetím odstavci na str. 21 se postupně mluví nejprve o tom, že data reprezentují K nezávislých výběrů z multinomického rozdělení (se čtyřmi kategoriemi reprezentovanými hodnotami vektoru (X, Y)). Ihned poté se však z tohoto stává K sad dvou nezávislých výběrů z alternativního rozdělení, tj. data reprezentují K nezávislých výběrů z podmíněných rozdělení $X \mid Y$.
- Hodnoty testové statistiky v simulační studii na str. 22, které software vykazuje jako NaN (*Not a Number*) by bylo vhodnější řešit příslušným limitním přechodem.

Navzdory uvedeným připomínkám však autor dle mého názoru předloženou bakalářskou prací prokazuje v dostatečné míře schopnost formulovat matematický text na úrovni vyžadované od absolventa oboru *Obecná matematika*.

Práce se zdroji. Zdroje, z nichž práce vychází, jsou řádně a též formálně správným způsobem citovány.

Formální úprava. Taktéž formální úprava práce je na dobré úrovni. Výskyt pravopisných chyb mohu, s ohledem na jazyk práce, kterým je slovenština, posoudit pouze v omezené míře. Nicméně zjevné překlepy nebo chyby se v práci vyskytují v míře zcela minimální a rozhodně nepřesahující míru obvyklou pro tento typ práce.

Závěr

Práci bez jakýchkoliv pochyb **doporučuji** uznat jako bakalářskou práci.

doc. RNDr. Arnošt Komárek, Ph.D.

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta Univerzity Karlovy

V Praze 10. srpna 2018