

## Posudek na bakalářskou práci Magdaleny Zvejškové Statistická chyba při reprezentativních výběrech z populace

Předložená bakalářská práce se zabývá problémem výpočtu statistické chyby a jejího zpřesnění ve výběrech z konečné populace a aplikací na odhady volebních preferencí. Práce je rozdělena do několika kapitol. V první jsou vysvětleny základní pojmy z výběrových šetření, ve druhé se autorka zabývá kvótním výběrem. Neuvažuje kvótní výběr jako nenáhodný, ale předpokládá, že kvóty jsou použity jen pro stanovení rozsahu skupin a respondenti jsou do skupin zařazováni náhodně, výběry do jednotlivých skupin jsou nezávislé. V tomto případě se vlastně uvažuje nezávislý multinomický model výběru známý z analýzy kategoriálních dat, přestože to není explicitně řečeno. Je zde dokázáno, že statistická chyba odhadu poměru kladných odpovědí v kvótním výběru, který uvažuje jednu nebo více kategoriálních proměnných, je menší než statistická chyba stejného odhadu v kvótním výběru, který kategorizaci neuvažuje. Ukazuje se, že poměr odhadů rozptylu (redukční koeficient) je doplněk do koeficientu determinace. Pro výběry s váženými daty již není podobný postup dokázán explicitně, jsou pouze provedeny simulace a empirické porovnání rozptylu váženého a neváženého odhadu.

Ve třetí kapitole jsou zpracovány průzkumy volebních preferencí hlavních politických stran pro volby do Poslanecké sněmovny parlamentu ČR v r. 2010 na základě dat získaných agenturami MEDIAN a STEM v letech 2006-2010. Časové řady volebních preferencí jsou nejprve vyrovnány polynomy a jsou analyzována rezidua. Potom jsou spočteny redukční koeficienty pro odhad statistické chyby, jednotlivě pro každou uvažovanou stranu a zvláště pro obě agentury. Podobný postup je aplikován také v případě, že místo eliminace trendu se uvažuje diferencování první řádu. Nakonec jsou porovnány výsledky obou agentur. Výsledky jsou prezentovány formou tabulek a grafů.

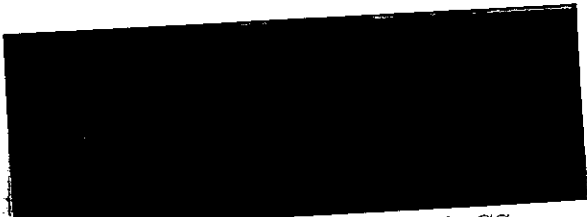
### Připomínky a dotazy

- Statistické pojmy jsou používány nepřesně, mnohdy není např. jasné, zda jde o rozptyl odhadu nebo odhad tohoto rozptylu (str. 13 a další), místo  $u(\alpha/2)$  má v daném kontextu být  $u(1-\alpha/2)$  (str. 8), na str. 29 a dalších se mluví o Spearmanově korelačním koeficientu, což není totéž co první autokorelační koeficient. Proč má být  $r(1) = 0$ ?
- Není doloženo, jakým způsobem bylo provedeno prokládání polynomů v časových řadách, ani jak byly získány hodnoty testových statistik a další statistické výpočty.
- Ve výpočtu ve vzorci  $\widetilde{\text{var } \hat{p}}$  na str. 13, 14, 15 je drobná chyba-mohla se projevit ve výpočtu koeficientu  $C$  v kap. 3?
- Seznam literatury je tvořen převážně odkazy na webové stránky, přestože literatura [2], [12] a [13] existuje v tištěné a recenzované verzi; odkazy na literaturu [4] a [5] nebyly na uvedených adresách nalezeny.

Většinu chyb lze vysvětlit malou zkušeností autorky; také relevantní odkazy na metodologii kvótních výběrů se vyhledávají obtížně. Na druhé straně bych ráda zdůraznila, že autorka

zpracovala téma z hlediska odborného i společenského velmi aktuální, prokázala schopnost samostatné práce, a proto doporučuji její práci jako bakalářskou práci na MFF UK v oboru Obecná matematika.

V Praze 11. června 2010



Doc. RNDr. Zuzana Prášková, CSc.  
recenzentka