

Bc. Pavel Poul: Ekologická regrese
(Posudek vedoucího)

Hodnocený text vznikl přepracováním neobhájené diplomové práce. Zabývá se problémem, který vzniká, když chceme popsat závislost panující na úrovni základních jednotek (např. osoby), ale k dispozici jsou pouze agregovaná data (např. pro územní celky). Zmíněný problém se vyskytuje v sociologii, epidemiologii nebo při zkoumání chování voličů. Literatura na toto téma je značně heterogenní, převládají texty z matematického pohledu nepřesně formulované. Úkolem diplomanta bylo problém popsat, vysvětlit a ukázat dostupné výpočetní možnosti.

Práce má v zásadě tři části. V první, nazvané *Důsledky agregace dat* se snaží popsat řešený problém jednak v rámci lineárního modelu, jednak jako model s náhodnými efekty, v obou případech za předpokladu normálního rozdělení. Tato část používá matematické formulace a zřejmě nemá konkrétní blízký vzor v časopisecké literatuře.

Další část nazvaná *Nula-jedničkové proměnné* je přehledem postupů navržených pro situaci, kdy závisle i nezávisle proměnná nabývají jen dvou hodnot.

Část nazvaná *Aplikace* ukazuje možnosti programového prostředí R k řešení úloh ekologické regrese. Jako výchozí data pro simulace použil diplomant údaje z amerického prostředí dostupné prostřednictvím R.

Práci, zejména kapitulu o nula-jedničkových veličinách, pojal diplomant zřejmě příliš široce, takže v čase, který vypracování diplomky věnoval, nedokázal dostatečně přesně a srozumitelně některé metody popsat, jednotlivá tvrzení přesně zformulovat a dokázat. Ne vždy je například zřejmé, zda je na daném místě textu nezávisle proměnná x či X chápána jako pevné číslo nebo jako náhodná veličina. Totéž se týká parametrů (či chybějících pozorování) β .

Některé konkrétní připomínky či dotazy:

1. Je na místě předpoklad o normálním rozdělení y_i na str. 23 v kapitole o nula-jedničkových proměnných?
2. Nerozumím výkladu o odhadu indikátorů D_i na str. 23. Vysvětlovaná proměnná v (2.33) přece není binární.
3. Ve výkladu kolem $CUSUM_t$ resp. $CUSUMQ_t$ mi schází kvantifikátor. Kdy zamítám nulovou hypotézu o neměnnosti?
4. Lze něco říci o spolehlivosti intervalu $[\theta^L, \theta^U]$ navrženém v (2.44) na str. 29?
5. Je možné rozdíl mezi dvěma odhady (tedy náhodnou veličinu) nazývat vychýlením? (str. 9, pod (1.15))
6. Co jsou u simulací (str. 44) *pozorované průměry*? Nebylo možno využít Tvrzení 2.11?

Diplomant pracoval až příliš samostatně. Shromáždil spoustu literatury psané na nejrůznější úrovni, používající nejrůznější značení. Popsal řadu modelů a metod řešení, jak byly postupně jednotlivými autory navrženy. Ukázal použití prostředků, které poskytuje statistické prostředí R. Na textu je bohužel patrný spěch, s jakým byla práce dokončena. Diplomový úkol byl splněn, ale s řadou nedostatků. S ohledem na množství práce, kterou do přípravy textu pan Poul vložil **doporučuji uznat předloženou práci jako práci diplomovou.**

V Praze dne 30. prosince 2015

Karel Zvára