

Posudek na diplomovou práci Pavla Poula

Ekologická regrese

Diplomová práce se zabývá regresní analýzou agregovaných dat a jejími důsledky pro neagregovaná data. Má 3 kapitoly, první je úvod do problematiky a zabývá se vztahy mezi parametry v modelech pro agregovaná data a parametry v regresních modelech pro původní neagregovaná data, ve druhé jsou představeny různé přístupy k analýze agregovaných dat, zejména binárních. V poslední kapitole jsou popsány metody porovnány na simulovaných a reálných datech. Na přiloženém CD jsou uloženy zdrojové kódy, seznamy použitých softwarových balíčků a data pro aplikační část práce.

Jde o přepracovanou verzi neúspěšně obhájené diplomové práce, jejíž teoretická část neodpovídala požadavkům na diplomovou práci. Autor podstatným způsobem změnil především první kapitolu a vynechal druhou. Další text je změněn již méně, částečně je změněn a rozšířen výklad o jednoduchém a sofistikovaném Goodmanově modelu, v partii o Kingově modelu byly opraveny nesprávné formulace a tiskové chyby a celkově se zlepšila jazyková úroveň práce. Nicméně některé partie by si i nyní zasloužily podrobnější výklad či komentář a objevily se i nové tiskové chyby a další nedostatky.

Připomínky:

- Str. 5, ř. 13 zdola: Symbolem $\mathbf{1}$ je značen vektor samých jedniček, ale ve výrazech $(\mathbf{1}, \mathbf{X})$, $(\mathbf{1}, \bar{\mathbf{X}})$ jde o vektory různé dimenze, podobně na dalších místech. Navíc se stejné značení používá jako indikátorová funkce, např. ve vztahu (2.54) na str. 33.
- Str. 9, vzorec (1.15): Ve výpočtu zřejmě má být (v souladu s (1.14) na str. 8) $\hat{\mathbf{b}}_b, \hat{\mathbf{b}}_w$. Co ale potom přesně znamená vychýlení v důsledku agregace a vychýlení v důsledku specifikace?
- Str. 9, ř. 13 zdola má být $Y_{it} = \xi_i + \varphi_{it}$
- Str. 14, příklad 2.1: Nebylo možné generovat hodnoty pravděpodobností přímo z intervalu $[0, 1]^2$, nikoliv je získávat zaokrouhlováním?
- Str. 16, vzorec (2.15): Správně má být $\hat{\beta}^0 = \bar{Y}_{\bullet\bullet} - \hat{p}\bar{x}_{\bullet\bullet}$
- Str. 21, vzorec (2.26): Je h_i totéž co $\bar{x}_{i\bullet}^2$?
- Str 22: Co je $\bar{j}_{i\bullet}$ ve vzorci (2.28)?
- Str. 23 poslední řádek: Proč právě \mathbf{X}_2 má mít plnou hodnotu?
- Str. 24, řádek 4: zřejmě má být \hat{Y}_t . Celý odstavec 2.6.1 je poněkud nejasný. Předpokládají se dodatečné proměnné \mathbf{z}_i ; do modelu vstupují ještě další proměnné D_i^* a další parametry γ , ty se předpokládají známé?
- Zcela nepochopitelná a nevysvětlitelná je struktura kapitoly 2. Po jakémsi úvodu následuje podkapitola 2.2, která se dělí na 2.2.1, 2.2.2a 2.6.1, poté následují podkapitoly 2.7, 2.8, 2.10 a 2.16 (která se dále dělí na 2.16.1 a 2.16.2). Tato zřejmě formální chyba značně snižuje kvalitu práce.