

Oponentský posudek na bakalářskou práci:

V. HERRMANN: FAVORITISM UNDER SOCIAL PRESSURE:
EVIDENCE FROM ENGLISH PREMIER LEAGUE

Předložená práce analyzuje data o nastaveném čase v zápasech anglické fotbalové ligy pomocí normálního lineárního modelu pro logaritmus délky prodloužení.

Teoretický základ lineární regrese je probrán ve druhé a třetí kapitole; zajímavý je zejména výklad týkající se kategorických vysvětlujících proměnných a korektní popis použití dummy proměnných.

Největším přínosem je analýza skutečných dat v kapitole 4, kde by však bylo vhodné doplnit alespoň stručné ověření předpokladů (např. grafy residuí).

Další připomínky:

Definition 1 Překlep v definici matice X . Splňuje $X_i = \varepsilon_i$ uvedené předpoklady?

str. 6, první odstavec v 2.1.1 Není mi jasné, jak se transformací závisle proměnné ztratí veškerá informace o vlivu regresorů.

str. 6, sekce 2.1.1 Prosím o doplnění výpočtu zvýšení střední hodnoty Y_i (viz “increase of the expected value of $\dots Y_i$ ” ve druhém odstavci).

str. 7 Zřejmě $A = B$ a není tedy nutné dvakrát opakovat stejný výpočet.

str. 8, věta 2 Opravdu jsou parametry uvedených normálních rozdělení náhodné veličiny (závislé na náhodné matici X)?

str. 8, první důsledek věty 2 Jak se v důkazu využije delta metoda?

str. 9, věta před větou 3 Proč je $X^{(q)}\beta^{(q)} = 0$ a $q = 0$?

str. 9, poslední odstavec V praxi se obvykle používá spíše test pro $q = 1$.

str. 12, poslední odstavec Interpretace uvedených hypotéz asi není správná. Stejný test lze provést i použitím prvního důsledku věty 2. Zcela korektní by bylo použití Tukeyho metody (tzv. studentizovaného rozpětí).

str. 13 Proč se parametry volí náhodně? Nebylo by jednodušší zvolit je pevně?

str. 14, první řádek Není level of significance spíš 5%?

str. 17, první odstavec Použité grafy jsou irelevantní, protože předpokládáme normalitu reziduí (tj. ε_i) a ne normalitu marginálního rozdělení závisle proměnné (tj. Y_i).

str. 17, 4.2.2 V praxi se vysvětlující proměnné obvykle odstraňují po jedné (kvůli možné závislosti, které ovlivňují sílu testů).

str. 24, 4.5.2 Stačilo by uvést výsledky testů, není nutné uvádět kompletní výstup pro stále stejný (pouze jinak parametrizovaný) lineární model.

Analýza dat o nastaveném čase je hezkým příkladem použití normálního lineárního modelu, přestože není zcela jasné, jestli jsou splněny všechny potřebné předpoklady (např. nezávislost a normalita). I přes drobné nedostatky doporučuji předloženou práci uznat jako práci bakalářskou.

V Praze 29. května 2017

Doc. RNDr. Zdeněk Hlávka, Ph.D.
KPMS MFF UK
Sokolovská 83, 186 75 Praha 8