

RNDr. Karel Hron, Ph.D.

Katedra matematické analýzy a aplikací matematiky
Přírodovědecká fakulta UP
17. listopadu 12, 771 46 Olomouc

Posudek oponenta diplomové práce
Bc. Petr Dokoupil: Aplikovaná logistická regrese

Diplomová práce studenta druhého ročníku oboru Aplikace matematiky v ekonomii má 97 stran, uvádí 11 literárních pramenů a je rozčleněna do devíti kapitol (plus příloha). První z nich (str. 7-10) je věnována zobecněnému lineárnímu modelu, druhá (str. 11-12) a třetí (str. 13-14) potom postupně rozdělení pravděpodobností exponenciálního typu, resp. linkové funkci v lineárních modelech. Ve čtvrté kapitole (str. 15-20) jsou shrnuty základní poznatky o standardní logistické regresi a následně v kapitole páté metody odhadu parametrů v logistickém modelu (str. 21-34). Příslušná statistická inference (testování hypotéz, intervalové odhady) je zmíněna v kapitole šesté (str. 35-38). Rozšíření modelu logistické regrese pro případ multinomické vysvětlující proměnné obsahuje sedmá kapitola (str. 39-52). V kapitole osmé (str. 53-63) je popsáno prostředí statistického softwaru SAS, který je použit při zpracování příkladů v praktické části neboli v deváté kapitole (str. 64-88) týkajících se výběru herní pozice pro hráče basketbalu a volby prezidenta.

Hlavním cílem práce bylo představit logistickou regresi jako oblíbený nástroj statistické analýzy kategoriálních dat a tento následně aplikovat na konkrétní reálné problémy týkající se sportovního managementu, resp. volby prezidenta ČR. Logistická regrese se přitom v poslední době stále častěji stává součástí základních kurzů matematické statistiky. Protože pro odhady parametrů v tomto regresním modelu neexistuje explicitní vyjádření, podobně jako je tomu v případě známé metody nejmenších čtverců, bylo například potřeba uvést příslušnou alternativu související s řešením soustavy rovnic v metodě maximální věrohodnosti a následně též přizpůsobit související statistickou inferenci. Řešení samotných reálných problémů si potom vyžádalo volbu vhodného modelu tak, aby byl tento současně jednoduchý a přitom zahrnoval všechny důležité vysvětlující proměnné.

Lze konstatovat, že uvedeného cíle bylo s menšími výhradami dosaženo. Práce je psána se snahou o systematizaci dostupných poznatků o daném tématu, oceňuji též originální volbu praktických problémů, k jejichž analýze byla logistická regrese využita. K práci mám přesto dvě metodické poznámky,

související s potřebou větší ucelenosti práce, a několik drobnějších připomínek. První kapitola není dle mého názoru dostatečně provázána s tématem práce, krátká zmínka je uvedena pouze na začátku páté kapitoly. Stejně tak asi nebylo nutné zmiňovat v kapitole 5.1 teorii maximálně věrohodných odhadů ve svém nejširším teoretickém rámci, stačilo rozebrat její aplikaci v případě studovaného modelu logistické regrese. Jako každé práci většího rozsahu, ani této se nevyhnuly menší překlipy a nepřesnosti:

- 13³: observační vektor nejprve označen $\mathbf{V}$ a následně bez bližšího vysvětlení jako $\mathbf{Y}$;
- 13₃: v jednom řádku značí $\mathbf{X}$ matici i vektor;
- 16₆: oba vektory, β a $\mathbf{X}$, by měly být sloupcové;
- 29⁸: v české literatuře obvykle rozdělujeme pojmy varianční a kovarianční matice, zde se jedná zřejmě o matici varianční;
- 31¹²: matice $\mathbf{W}^{1/2}$ není definována;
- 42₁: užívaná procedura LOGISTIC je představena teprve v osmé kapitole; obecně, software SAS stálo za to představit daleko dříve (někde před str. 28) a zdůvodnit výhodnost jeho použití před ostatními alternativami (též námět do obhajoby);
- nestandardní formát citací, zvláště při použití anglické spojky "and" (za všechny 31₃);
- chybí důsledné užívání italiky pro označení matematických objektů a jejich indexů (za všechny 60¹, 60¹⁰, 60¹²);
- 76⁴: formulace, místo "pro hladinu testu" raději "na hladině testu";
- 86₃: jestliže nejsou uvedené vysvětlující proměnné (potažmo příslušné regresní parametry) signifikantní, nabízí se přirozeně otázka, zda nebylo lepší zvolit nějaké jiné?
- drobné překlipy: 31₁₀, 44₉, 75¹⁰, 80₉, 82 (popisy obrázků), 86₃.

Vzhledem k výše uvedeným připomínkám navrhuji do obhajoby hodnocení C .

V Olomouci, dne 30. dubna 2012