

Posudek oponenta na diplomovou práci Bc. Markéty Daněčkové
Výběr proměnných v regresi

Práce se věnuje tématu, které často tvoří součást praktického zpracování dat pomocí regresní analýzy. Je totiž třeba rozhodnout, které proměnné do modelu zařadit a které naopak nezařadit, a nějakým způsobem provést srovnání perspektivních modelů. Je známo, že zautomatizované krokové procedury zařazování/vylučování proměnných mohou být zdlouhavé, a ani tak není vždy zaručen uspokojivý výsledek – v praxi je nejlépe kombinovat je se subjektivní úvahou podloženou zkušeností.

Teoretická část práce pokrývá především kritéria pro porovnání modelů (kapitola 3), a sice kritéria odvozená z věrohodnostní funkce, která používají penalizaci za zahrnuté parametry. Zde diplomantka spočítala 6 příkladů na výpočet K-L divergence: ručně i s využitím softwaru. Další vodítka pro posuzování kvality modelů poskytuje analýza reziduí (kapitola 4). I zde jsou uvedeny příklady, formou komentovaných výstupů z balíku R. Praktická část (kapitola 5) řeší problém výstavby vhodného modelu na dvou skutečných datových sadách: procento tělesného tuku a spotřeba benzínu.

Práce je psaná dobře čtivým slohem, má přehledné členění a pěknou grafickou úpravu, včetně ilustrativních tabulek a grafů. Teorie je prezentována srozumitelně, značení je jednotné. Příklady jsou zpracovány dostatečně podrobně a vhodně komentovány. Po obsahové a věcné stránce jsem nenašel žádné zásadní nedostatky.

Několik mých výhrad je ke gramatice:

- y ve 4. pádě: Leiblerovy divergence, Fisherovy matice,
- mn. č. stf. rodu: ta kritéria, ta data, ta pozorování (nikoli ty),
- i v min. čase sloves: (my) jsme udělali (nikoli –ly),
- 3. os. mn. č. nacházejí (nikoli nachází),
- přivlastňovací tvar: Quinnovo IC,
- adjustované R^2 (nikoli adjungované).

Jinak se v textu vyskytují překlepy a chybné větné čárky, není jich však příliš mnoho.

Několikrát se v jednom vzorci objevuje značka pro logaritmus zapsaná dvěma různými fonty (str. 32, 43, 46). Při obhajobě bych rád slyšel odpověď na tyto otázky a připomínky:

- str. 30 ... Ruční výpočet označil za bližší model jiný než software. Čemu byste dala přednost a proč?
- str. 44₅ ... Proč lze vynechat poslední dva výrazy?
- str. 45₄ ... Vysvětlete, proč *BIC* vybírá model s nejmenším *K*.
- str. 55 ... Mám za to, že v druhém vzorci pro D_i má být standardizované (normované) reziduum e_i^* , nikoli studentizované (jackknife) reziduum e_{-i}^* .
- str. 62 ... Lépe vysvětlete, jak souvisí volba *AIC* vs. AIC_C s metodou výběru proměnných.

Celkově hodnotím práci velmi pozitivně, diplomantku je třeba pochválit za řadu věcí: za zpracování textu, za ruční výpočty, za práci se softwarem R, za přehlednou prezentaci výsledků. Diplomovou práci doporučuji k obhajobě a navrhuji hodnocení **A**.

V Šumperku 15.1. 2014

RNDr. PhDr. Ivo Müller, Ph.D.
oponent