

Posudek oponenta na diplomovou práci Bc. Jana Strnada „Analýza storna pojistných smluv“

Práce se zabývá využitím logistické regrese k analýze stornovosti smluv pojištění odpovědnosti z provozu motorových vozidel.

Těžiště práce spočívá v analýze reálných dat vedoucí k sestavení modelu, umožňujícího na základě vybraných proměnných odhad pravděpodobnosti storna z důvodu nezaplacení pojistného při obnově smlouvy.

Z praktické části práce je patrné, že student věnoval značné úsilí tvorbě modelu s využitím statistického softwaru SAS. Postup tvorby a validace modelu i jeho kalibrace s ohledem na trend ve vývoji stornovosti je v práci podrobně popsán a doplněn množstvím ilustrativních tabulek a grafů. Tato část textu je psána kultivovaně a srozumitelně, jedná se však spíše o „povídání“ určené čtenáři obeznámenému s příslušnými statistickými procedurami a testy.

Matematický popis těchto metod v teoretické druhé kapitole má výrazně podprůměrnou úroveň a vzbuzuje pochybnosti, zda autor matematické stránce tématu dobře rozumí.

Uvádím zde některé konkrétní připomínky a nejasnosti:

- Poslední dvě věty v předposledním odstavci str. 8 nejsou pravdivé.
- První vzorec na str. 13 je vyjádřením vektoru parciálních derivací (gradientu), formulace, kterou je uveden, je nepřesná.
- Str. 13, ř. 4: „položením těchto rovnic rovných nule“ - nesmyslná formulace.
- Str. 14, ř. 5: Má testová statistika (2.2) za uvedené hypotézy přesně χ^2 -rozdělení?
- Str. 14 - zcestný popis Waldova testu v odst. 2.8: „Waldův test porovnává (??) maximálně věrohodné odhady regresních koeficientů daných proměnných s odhadem jejich směrodatných odchylek.“ Dále - co je testová statistika, o které se zde tvrdí, že má normální rozdělení? Odkdy má druhá mocnina normálně rozdělené veličiny χ^2 -rozdělení s p stupni volnosti?
- Str. 14, odst. 2.9: Co znamená „je porovnávána“ na 6. řádku zdola? Jak z tohoto vyplývá tvar testové statistiky? Co je „počet omezení daných hypotézou“?
- Str. 17: Ve vzorci (2.3) se vyskytuje statistika X^2 , která je definována až na následující straně.
- Str. 18, ř. 14 zdola: „podmíněné střední hodnoty náhodné veličiny Y_i “ - čím je podmíněná?
- Str. 22, ř. 4: Jak se má rozumět formulaci „Protože budu pracovat se skutečnými daty, jsem při jejich zpracování omezen pouze na tento statistický program.“ (míněn SAS)?

- Odkud pocházejí data využitá v práci?

K formální stránce práce mám následující připomínky:

- V odborných textech není obvyklé psát v 1. osobě jednotného čísla.
- Vzorce jsou většinou nečíslované a mnohdy uvozeny pouze dvojtečkou, takže práce netvoří souvislý text.
- Za nedostatečné považuji citace literatury, zejména není zřejmé, odkud jsou čerpány informace v odstavcích 2.5 - 2.9. Proč jsou základní definice typu Giniho koeficientu a Lorenzovy křivky čerpány z prezentace k přednášce umístěné na internetu?
- Práce obsahuje i řadu gramatických chyb, zejména chybějících čárek v souvětích, ale i několik chyb v pravopisu („schoda“).

Práce po matematické stránce vykazuje závažné nedostatky. Výsledky analýzy dat v ní provedené však podle mého názoru mohou být přínosem pro praxi pojišťoven. Práci proto doporučuji uznat za diplomovou.

V Praze 10.9.2013

RNDr. Lucie Mazurová, Ph.D.