

Posudek na bakalářskou práci Petry Kynčlové,

studijní obor Matematika – ekonomie se zaměřením na pojišťovnictví,
KMAaAM PŘF UP Olomouc, školní rok 2009/2010

Statistická analýza dojivosti v programu SAS

Bakalářská práce Petry Kynčlové na téma Statistická analýza dojivosti v programu SAS má celkem 44 stran a je rozčleněna do šesti kapitol.

V první kapitole je čtenář stručně seznámen s prostředím modulu SAS Enterprise Guide (SAS EG), tzv. tenkého klienta SASu, který umožňuje přehledným způsobem manipulovat s daty, provádět různé (nejen) statistické analýzy a prohlížet výstupy, které mohou dále sloužit jako vstupy dalších procedur. Autorka popisuje prostředí jako přehledné a intuitivní, proto se podrobnějším popisem jednotlivých úloh (tasks) nezabývá.

Před samotnými analýzami, je ve druhé kapitole představena struktura souboru s daty, který je natolik rozsáhlý, že by jej nebylo vhodné vkládat do práce ani jako přílohu.

Autorka analyzuje data o dojivosti nejprve pro všechny kravíny, a to ve třetí kapitole. Používá srovnávací krabicový graf a srovnávací histogram jako grafickou metodu popisné statistiky reprezentující názorně distribuci dat - polohu, variabilitu, symetrii a částečně podobnost s normálním rozdělením pravděpodobnosti. Následuje příklad testování hypotézy o shodě středních hodnot dojivosti v kravíně Kunín a Hradecká a.s. prostřednictvím dvouvýběrového t-testu. Předpoklad normality je ověřován graficky využitím normálního kvantilového grafu. Na základě získaných výsledků t-testu zamítá nulovou hypotézu o shodě středních hodnot dojivosti v těchto dvou vybraných kravínech.

Pro porovnání středních hodnot dojivosti všech kravínů, jejichž údaje jsou k dispozici, chce autorka použít analýzu rozptylu. Bohužel zjišťuje, že je porušen předpoklad homoskedasticity (výsledek Bartlettova testu), proto volí neparametrickou obdobu testu ANOVA, a to Kruskalův-Wallisův test. Na základě výsledků tohoto testu dochází k podobnému závěru jako v předchozím případě, tj. že se dojivost v jednotlivých kravínech liší. Které dvojice tuto odlišnost způsobují odhaluje prostřednictvím Scheffého metody násobného porovnávání. V závěru kapitoly jsou také popsány dva testy normality založené na výběrových momentech – výběrové šikmosti a špičatosti. Také tyto testy odhalují, že je porušen další z předpokladů použití testu ANOVA – normalita dat, a tedy použití neparametrického Kruskalova-Wallisova testu bylo správné.

Další zajímavou analýzou je porovnání dojivosti v závislosti na laktačním cyklu pro data z jednoho kravína (Kunín), kterou autorka rozpracovala ve čtvrté kapitole práce. Podobně jako v předchozím případě testuje nejprve předpoklad homoskedasticity dat Bartlettovým testem a na základě jeho výsledku volí opět neparametrický Kruskalův-Wallisův test. I v tomto případě zjišťuje, že dojivost je statisticky významným způsobem ovlivněna faktorovou proměnnou, tj. laktačním cyklem. V závěru kapitoly opět provádí násobné porovnávání a graficky testuje normalitu dat prostřednictvím normálních kvantilových grafů.

Zatímco předchozí kapitoly byly věnovány porovnávání skupin dat, v kapitole páté se setkáváme s regresní analýzou. Autorka inspirovaná publikovanými výsledky, odhaduje závislost dojivosti na dnu laktace pomocí kubické regresní křivky. Neméně zajímavá je i poslední kapitola práce zabývající se analýzou přežívání, tj. doby do vyřazení dojnice z chovu.

Práce je psána přehledně s uvedením vybraných teoretických poznatků v úvodu ke každé použité metodě. Výsledky jednotlivých procedur prováděných v prostředí SAS EG jsou uváděny převážně ve formátu tabulek a grafů převzatých přímo z výstupů tohoto programu. Čtenář si tak dokáže udělat představu o tom, jaké výstupy a v jakém formátu SAS EG produkuje. Záběr práce je skutečně široký, a tak není divu, že se již nedostalo prostoru pro podrobnější vysvětlení použitých procedur SAS EG. To ale nebylo ani cílem práce. Naopak, autorka si kladla za cíl pouze využít SAS pro analýzu konkrétních dat a tento záměr se jí podařilo naplnit.

K práci mám tyto drobné připomínky a otázky:

- str. 7₇ Překlep – „diogově“.
- str. 10 Dvojí formátování střední hodnoty.
- str. 10_{6,7} Překlep – „kvantil“.
- str. 11_{3,4} Označení centrálních momentů.
- str. 12 Krabicový graf neodpovídá popisu v textu (typ SKELETAL, SCHEMATIC).
- str. 12₁ Chybí odkaz na obr.
- str. 11₂ Chybí relační operátor.
- str. 17₅ Překlep v symbolu sumace.
- str. 17₈ Opačný relační operátor.
- str. 18 Chybí mi vysvětlení, který řádek z tabulky t-testu použijeme a proč.
- str. 21 Chybí popis významu použitých symbolů Y_{ij} , y_i .
- str. 23₃ Formátování indexu.
- str. 36₇ Překlep v mocnině prvku matice $\mathbf{X}$.
- str. 42 Vysvětlit význam osy y v grafu.

Výše uvedené připomínky nejsou závažné povahy a řadí se spíše do kategorie překlepů. Nesnižují kvalitu této široce a prakticky zaměřené práce, proto doporučuji tuto bakalářskou práci k obhajobě a v případě uspokojivého vysvětlení některých nejasností (viz výše) u obhajoby navrhuji hodnocení stupněm **výborně**.

V Olomouci dne 9. 5. 2010

Mgr. Jana Vrbková
oponent