

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Analysis of fluctuation of labourers

Autor: Ondřej Zeman

SHRNUTÍ OBSAHU PRÁCE

Cílem práce bylo řešení praktického problému a to statistická analýza fluktuace zaměstnanců blíže nespecifikovaného („anonymního“) výrobního podniku s provozy v Mladé Boleslavi, Kvasínách a Vrchlabí. Úkol byl řešen v úzké spolupráci s pracovníky tohoto podniku. Práce je členěna do tří kapitol doplněných o úvod a závěr. První kapitola poskytuje obecný úvod do problematiky fluktuace pracovníků a seznamuje čtenáře se strukturou dostupných dat (poskytnutých výrobním podnikem). Jsou zde diskutovány též nedostatky těchto dat a navrhovány postupy, jak se s těmito nedostatky vyrovnat. Druhá kapitola shrnuje přehledovou formou matematicko-statistický aparát, který je následně použit k řešení zadaného problému. Třetí (nejobsáhlejší) kapitola seznamuje čtenáře s výsledky analýzy reálných dat a ve své podstatě řeší zadaný problém.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Jedná se o jednu z prvních diplomových prací obhajovaných na oboru *Pravděpodobnost, matematická statistika a ekonometrie* na MFF UK zaměřených na řešení praktického problému ve spolupráci s průmyslem. Autor tedy neměl k dispozici prakticky žádné starší práce obdobného zaměření, v nichž by se mohl inspirovat ohledně formy textu a dalších náležitostí. Tento aspekt považuji na tématu práce za nejobtížnější a autor se s ním dle mého názoru popasoval se ctí. K uspokojivému vyřešení zadaného problému bylo též nutné sáhnout k metodám, jež nejsou standardně vyučovány v rámci navazujícího magisterského studia (a autor tak učinil). Lze tedy konstatovat, že zadání práce přesahovalo rámec „rutinní“ analýzy dat (což by jistě měl být základní požadavek na práce tohoto typu). Zadání považuji předloženou prací v každém případě za splněné.

Vlastní příspěvek. Neoddiskutovatelným vlastním příspěvkem autora je netriviální analýza reálných dat a prezentace jejich výsledků ve třetí kapitole. Též shrnutí možných přístupů v rámci druhé kapitoly, byť je přehledového charakteru, lze v této podobě považovat za vlastní příspěvek autora.

Matematická úroveň. Čistě matematická část je obsažena pouze v rámci přehledové druhé kapitoly, což ale s ohledem na povahu zadání nepovažuji za nedostatek, neboť těžiště práce, měl-li být splněn diplomový úkol, leželo jinde než ve formulování a dokazování matematických tvrzení. Drobné připomínky ke značení a matematickému vyjadřování jsou následující:

1. V definici 1 na str. 9 je předmětná náhodná veličina odpovídající odezvě značena jako X . Toto je pro čtenáře matoucí, neboť hned na další straně (a dále) je náhodná veličina se stejným kontextem značena jako Y a písmeno X přebírá roli vysvětlujících proměnných.
2. Na str. 10 uprostřed se mluví o alternativním rozdělení, jsou uváděny vzorce pro kanonický parametr a další související veličiny. Ve vzorcích se vyskytuje písmeno p , jehož význam si čtenář musí domyslet. Co horšího, o několik řádek níže začíná p znamenat počet vysvětlujících proměnných.
3. Vyjádřením „... \mathbf{Y}_i is an independent random vector“ (str. 12 nahoře) se patrně myslí, že \mathbf{Y}_i je vektorem s nezávislými složkami.
4. Uprostřed str. 23 se tvrdí „... the Akaike information criterion (AIC) can be used for choosing the best regressors in the model.“ S tímto bych si dovolil nesouhlasit. AIC (a další informační kritéria) lze použít nejvýše jako pomocné ukazatele, má-li být vybudován smysluplný regresní model.

5. Význam symbolu D vzorci (2.31) na str. 23 není zaveden.

Zásadnější výhradu mám potom k zavedení modelu pro multinomickou odezvu v oddíle 2.5. Model je zaveden dosti polopaticky a musím konstatovat, že pokud bych odjinud nevěděl, o co se jedná, z diplomové práce bych to spíše nepochopil. K nepochopení (kromě jiného) přispívá vyjádření "*It holds that...*" před výrazem (2.37) na str. 25. Domnívám se, že výraz (2.37) vyjadřuje předpokládaný model, kterým by bylo záhodno začít před tím, než se začne tvrdit, že *něco platí*.

Faktická správnost. V práci se dle mého názoru nevyskytují zásadní faktické chyby. Na druhou stranu jsem v pouze omezené míře schopen posoudit na základě předloženého textu, zda byla praktická analýza provedena korektně (viz „zásadnější koncepční připomínka“ níže).

Formální úprava a práce se zdroji. Práce je psána anglicky, v některých pasážích stylisticky mírně neobratně, ale vzhledem k tomu, že se patrně jedná o autorův první delší anglicky psaný odborný text, lze konstatovat, že práce je po jazykové stránce vyhovující. Nalezneme zde i několik překlepů a jiných nedostatků (např. chybějící křížový odkaz na tabulku na str. 61), kterým by bylo možné se při větší pečlivosti vyhnout, ale celkově lze říci, že tyto nedostatky jsou přiměřené rozsahu a povaze práce. Zdroje, z nichž práce vychází, jsou řádně citovány. Autorovi bych nicméně vytknul použitý styl pro seznam literatury. Zdroje jsou správně a rozumně řazeny dle příjmení autorů, v seznamu literatury jsou však na prvním místě uváděna plná křestní jména. Tímto se čtenáři do značné míry ztěžuje nalezení požadovaného odkazu. Dále mě překvapuje pouze iniciála křestního jména autora třetího odkazu. Nebylo možné zjistit plné křestní jméno?

Další drobná poznámka k formální úpravě: matematické znaménko mínus (–) je až na výjimky vysázeno jako krátká pomlka (-).

ZÁSADNĚJŠÍ KONCEPČNÍ PŘIPOMÍNKA

K práci bych si dovilil jednu zásadnější koncepční připomínku, která by mohla sloužit též jako vodítko pro autory tématicky podobně orientovaných diplomových prací v budoucnosti. Kladně hodnotím druhou kapitolu, jež souvislým způsobem shrnuje aparát potřebný k analýze dat včetně faktu, že aparát jde nad rámec obvykle probírané látky. Méně kladně však hodnotím třetí „praktickou“ kapitolu, která je psána do jisté míry románovým stylem. Najdeme zde na jedné straně nepodstatné technické detaily ohledně předzpracování dat (např. „*We changed the value of staying in the company from 0 to 3, because of the easier manipulation with the data in the software*“ na str. 32, ale i jinde), na druhé straně zde některé zásadnější části chybí (viz dále). I některé podstatné aspekty v textu obsažené se ztrácejí v záplavě okolního balastu. Významná část textu se věnuje interpretacím čísel z mnoha obsáhlých tabulek, což může být nepochybně zajímavé pro zadavatele analýzy, ale je již méně zajímavé pro osobu, jenž má hodnotit diplomovou práci z pohledu její relevantnosti za účelem přiznání titulu Mgr. v příslušném oboru. I zadavatel bude nicméně postrádat shrnutí toho nejdůležitějšího, co si z výsledků odnést pro praxi (důsledky pro praxi je nutné „vyzobat“ z mnoha stran textu). Jako oponent potom postrádám o něco více informací týkajících se toho, jak se k jednotlivým modelům došlo, jak a zda byla vyhodnocena jejich relevantnost pro daná data atp. Oddíl 3.2 sice popisuje obecnou strategii použitou k vývoji jednotlivých modelů, ale jakákoliv diagnostika chybí. Moje osobní doporučení pro obdobný typ prací zní:

1. Vyprodukovat zprávu pro zadavatele, která se soustředí na interpretaci výsledků a jejich důsledky pro praxi. Zpráva bude obsahovat též shrnutí nejdůležitějších zjištění a bude sloužit jako (elektronický) appendix diplomové práce.
2. Do samotné diplomové práce (obdob kapitoly 3 předložené práce) zařadit zejména pasáže, jež odbornému recenzentovi umožní posoudit, zda byla statistická analýza provedena smysluplně a bez zásadních nedostatků. Ze samotných konkrétních výsledků potom vybrat pouze to nejdůležitější

a nejzajímavější. Diplomová práce by se v každém případě měla vyhnout sáhodlouhému popisu čísel z nekonečné sady tabulek.

DOPORUČENÍ PRO OBHAJOBU

1. Prosím, aby netriviální část obhajoby byla věnována zdůvodnění, že alespoň jeden z prezentovaných modelů (např. “*core model*”) je relevantní pro daná data. Zdůvodnění by přitom mělo přesáhnout postup, kdy se srovnávají hodnoty informačních kritérií.
2. Obhajoba by též měla obsahovat shrnutí nejdůležitějších zjištění z pohledu zadavatele (výrobního podniku).

ZÁVĚR

I přes výše uvedené výhrady se autor se zadaným úkolem popasoval dle mého názoru úspěšně a práci **doporučuji** uznat jako diplomovou pro obor *Pravděpodobnost, matematická statistika a ekonometrie* na MFF UK.

doc. RNDr. Arnošt Komárek, Ph.D.

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta Univerzity Karlovy

V Kačlehách 28. srpna 2020