

Hodnocení vedoucího diplomové práce

Autor/Autorka	Bc. Ivana Gabrišková
Název práce	Modelování a odhadování výsledků hokejových utkání
Studijní obor	Finanční informatika a statistika
Vedoucí práce	Ing. Patrice Marek, Ph.D.

Splnění cílů práce:

nadstandardně velmi dobře splněny s výhradami nebyly splněny

Odborný přínos práce:

nové výsledky netradiční postupy zpracování výsledků z různých zdrojů shrnutí výsledků z různých zdrojů bez přínosu

Matematická (odborná) úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Věcné chyby:

téměř žádné vzhledem k rozsahu přiměřený počet méně podstatné, větší množství podstatnější, větší množství závažné

Grafická, jazyková a formální úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Přístup autora k řešení práce, spolupráce s vedoucím práce:

samostatná práce s výbornou komunikací pečlivá práce, drobné zásahy vedoucího pečlivá práce, podstatnější zásahy horší komunikace špatný přístup k práci

Slovní hodnocení

Předložená diplomová práce se zabývá tvorbou a optimalizací modelů pro odhadování pravděpodobností výsledků v hokejových zápasech. Autorka čerpá modely z aktuální odborné literatury a navrhuje jednu vlastní úpravu modelů, která spočívá v individualizaci parametru domácí výhody, který je běžně uvažován jako globální. Po popisu dat a základní teorie jsou testovány obvyklé předpoklady modelů – Poissonovo rozdělení počtu vstřelených gólů a případná nezávislost mezi vstřelenými góly. Pátá kapitola práce se zabývá představením a odhadováním parametrů Maherova modelu, šestá kapitola se zabývá Dixon-Colesovým modelem a sedmá kapitola představuje vlastní úpravu modelů. Srovnání modelů je ukázáno v osmé kapitole.

Práce obsahuje zajištění, kontrolu a zpracování dat velkého rozsahu. U modelů pak autorka musela nastudovat některé modely z odborné literatury. Přístup autorky k vlastnímu zpracování diplomové práci musím ale hodnotit negativně, jelikož značná část práce byla ponechána na poslední týden a odevzdání tedy proběhlo velice hekticky. V práci proto zůstaly neopravené i některé mnou komentované chyby. V případě Maherova modelu bylo navíc potřeba mnoho konzultací, než se autorce povedlo vytvořit fungující model. Autorka nakonec všechny cíle práce splnila, ale práce je díky chybám a špatným, případně chybějícím, popisům spíše podprůměrná. Vytvořená makra by bylo také potřeba lépe optimalizovat, aby bylo odhadování urychleno, např. přímým předáváním hodnot. V některých částech také chybí motivace, proč autorka danou věc

zpracovává a jaké z toho plynou závěry – především v části, která se věnuje sázení, kde chybí u náhodného sázení motivace i závěry. Grafickou, jazykovou a formální úroveň práce lze hodnotit ještě jako průměrnou.

Podrobné připomínky k práci

Abstrakt

- Nemá být uváděn odkaz na literaturu.

1 Úvod

- Místo „porovnat“ má být „porovnání“ (odst. 1, str. 1).

2 DATA

- Název je, na rozdíl od ostatních kapitol, psán velkými písmeny.
- V práci se často místo pomlčky používá spojovník, např. str. 4 a pátý bod seznamu, kde navíc jedna pomlčka chybí.

3 Statistické pojmy a metody

- V části 3.2 chybí předpoklad k uvedené definici. Dále není vysvětlen symbol H_0 (str. 5).
- V části 3.3 je chybně uvedeno, že testujeme hypotézu, že „rozdělení veličiny X má nějaké rozdělení“ (str. 5).
- Podmínka, že očekávané četnosti musí být větší než 1, je nedostatečná, je potřeba připojit i požadavek, kolik jich může být menších než 5 (str. 6).
- U Bonferroniho korekce (str. 6) není uvedeno, proč je nutné upravovat hladinu významnosti. Navíc poslední věta v této kapitole nedává smysl.

4 Testování předpokladů modelů

- V prvním odstavci nedávají některé věty smysl (str. 8).
- Předposlední odstavec na str. 8 měl být uveden dříve.
- Závěry v tabulkách 3, 4 a 5 (str. 9 a 10) nesouhlasí s výše zavedeným značením.

5 Maherovy modely

- Je zde používán dvojitý formát značení (neplatí pouze v této části, ale obecně). Příkladem je λ_{ij} v textu a ve vzorci na str. 13. Navíc je symbol používán i bez indexů.
- Některé vzorce na str. 14 a 15 jsou napsány větším písmem než ostatní.
- Místo „0.83 krát“ má být použito „0.83krát“ (str. 16), toto se opakuje v celé práci.
- Není potřeba používat kvadrát odchylek, ale – vzhledem k použití – je vhodnější měřit absolutní chybu (str. 16). Dále není vhodné k číslu udávat poznámku pod čarou označenou číslem, nelze rozlišit, že se nejedná o druhou mocninu, navíc s uvedeným kontextem se nabízí na tomto místě druhou mocninu uvažovat.
- Na str. 20 je uvedeno, že ostatní sezony jsou zpracovány v excelu. V práci přitom úplně chybí okomentování a uvedení, jaké jsou výsledky a co z nich plyne.
- Při zhodnocení Maherova modelu je uvedeno, že je pro Extraligu nevhodný, toto je ale konstatováno pouze na základě jedné sezóny (str. 21).

6 Dixon-Colesovy modely

- Kapitola, na rozdíl od ostatních, nezačíná na nové stránce.
- Parametr β nevyjadřuje sílu obrany, jak uvádí autorka v práci (např. str. 22), ale slabost obrany.
- Parametr p nevyjadřuje pravděpodobnost remízy, jde o parametr používaný ke smíšení dvou rozdělení (str. 23).
- Funkce ve vzorci (6.12) již není věrohodnostní, ale pseudověrohodnostní, stejně tak (6.13) a (6.14) na str. 25.
- V textu na str. 27 je uvedeno, že od šestého kola jsou rozdíly odhadovaných parametrů menší než 0.2, obr. 5 ale ukazuje, že tomu tak není.
- Obr. 7 (str. 28) nepovažují dostatečný pro ilustraci extrému, bylo by vhodné zkusit i větší hodnoty parametru ξ .
- V druhém odstavci na str. 30 je uvedeno, že nezápornost parametrů vyplývá z logaritmické věrohodnostní funkce, což není pravda. Odkaz na funkci je navíc chybný.

- Vzorec na úvodu str. 33 není označen a p ani p_v nejsou pravděpodobnosti, jak je uvedeno v textu. Formát textu u vysvětlení parametru v je nevhodný.
- Na str. 33 se několikrát zmiňuje, že jsou uvedeny pravděpodobnosti. Jedná se ale pouze o jejich odhady.
- Na str. 39 je zbytečně vynecháno místo na konci stránky, toto se v textu dále opakuje vícekrát. Autorka tedy nevhodně umísťovala tabulky a obrázky.

8 Předvídací schopnost a sázeční strategie

- Chybí motivace pro použití náhodného sázení, tj. co tím chce autorka ukázat. V návaznosti na to chybí i vyhodnocení závěrů.
- Autorka zmíní velikost vstupního kapitálu až po tom, co prezentuje výsledky sázení (str. 60).
- Relativní hodnota majetku v tabulce 35, textu i následujících tabulkách a obrázcích je uvedena špatně. Má se vztahovat k počátečnímu majetku.
- Na str. 63 autorka uvádí intervaly ziskovosti pro parametr L . To ale nelze tvrdit v celých uvedených intervalech, jelikož byl parametr zkoušen jen pro vybrané hodnoty.
- Na str. 71 v posledním odstavci není jasné, pro jakou hodnotu parametru L je výsledek platný.
- Jak již bylo uvedeno, v celé kapitole 8.4 (str. 72) a 8.5 (str. 75) chybí motivace, proč se porovnávají výsledky s náhodným sázením, jaký to má smysl a jaké jsou výsledky.

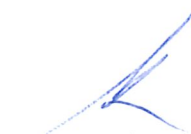
9 Závěr

- V pátém odstavci na str. 78 autorka srovnává modely dle $S(\xi)$, což není vzhledem k rozdílnému počtu zápasů možné.
- V předposledním odstavci na str. 78 nedává první věta smysl.

Otázky k obhajobě

1. Jaký byl smysl použití náhodného sázení a jaké z toho lze vyvodit konkrétní závěry? Ilustrujte na české Extralize.
2. Pokud byste měla srovnat modely všech tří lig na základě $S(\xi)$, co by k tomu bylo potřeba provést?

Navrhuji hodnocení známkou: **DOBŘE**



Patrice Marek, 4. 6. 2018

