

Prof. RNDr. Eduard Stehlík, CSc.,
Pplk. Sochora 15,
170 00 Praha 7

Oponentský posudek disertační práce

Statistické metody pro hodnocení predikční validity

Autorka: RNDr. Jana Rubešová

111 stran, 20 obrázků, 10 tabulek, CD s texty použitých skriptů

Jedním z důležitých problémů souvisejících s vysokoškolským studiem je otázka přijímání uchazečů do prvního ročníku. Vedle vážných zájemců o studium daného oboru se hlásí i maturanti, kterým zápis na vysokou školu umožňuje využívat např. sociálních výhod studentů a při tom i pracovat. Přijetí takových uchazečů může ztížit přijetí skutečných zájemců o daný obor. Ideální by proto bylo přijmout všechny přihlášené maturanty a v průběhu několika semestrů oddělit náročnými zkouškami a sledováním práce ve cvičeních a seminářích „zrno od plev“ tak, aby např. do 3. semestru se zapisovali skuteční zájemci ochotní věnovat se studiu.

To je ovšem ideál, který nelze realizovat. Překážkou jsou omezené kapacity fakulty, předpokládaný počet přijatých uchazečů, jehož překročení o určité procento je penalizováno. Proto je nutné provádět přijímací řízení, při němž se z přihlášených maturantů vybere příslušný počet šťastlivců, kterým je umožněn zápis.

Hlavní částí přijímacího řízení je obvykle zkouška z několika profilujících předmětů vyučovaných na střední škole, případně test „předpokladů ke studiu“ sestavený některou - organizací, většinou komerční. Je proto důležité zhodnotit, jaká je vazba mezi výsledkem přijímací zkoušky, prospěchu na střední škole, příp. jinými faktory, a úspěšností studia na fakultě a podle výsledků analýzy upravit přijímací řízení. Právě takovému hodnocení je věnována předložená disertační práce.

Práce obsahuje vedle úvodu dvě rozsáhlé kapitoly, závěr a přílohy A, B, C.

První kapitola je věnována přehledu použitých statistických metod a modelů od běžně používaných (korelační analýza, lineární regresní model) až po hodnocení shody predikce dvěma modely založenými na stejných datech. Autorka si v této kapitole stanovila obtížný a nevděčný úkol - popsat alespoň základní pojmy, postupy a vzorce ke všem statistickým metodám, které ve své práci použila. Uvést takový přehled ve stručné přehledné formě není jednoduché. Jde to celkem snadno u běžných na odvození nepříliš náročných postupů. Při uvádění speciálních metod je nutný odkaz na literaturu a odvození jenom naznačit. Užitečné je, že autorka u každé metody uvádí numerický příklad na základě dat týkajících se 140 studentů bakalářského oboru Geografie-kartografie zapsaných na fakultu ve školním roce 2003/04. Příklady jednak ilustrují způsob výpočtu, jednak ukazují, jak hodnotit výsledky.

Značnou pozornost v přehledu metod v kap. 1 věnuje autorka tzv. ROC analýze. Popis této metody ve zhuštěné formě není právě snadný, proto jsou některé pojmy jen uvedeny,

některá tvrzení jsou odvozena. Výklad navazuje na předcházející odstavce 1.3.1 (Logistická regrese), který jsem bohužel nemohl sledovat, neboť v mém výtisku práce chybí stránky 30-32. ROC křivky a pojmy s tím spojené jsou ve 2. kapitole užity k hodnocení predikce úspěšnosti studia za různých kombinací výchozích prediktorů.

Ve druhé kapitole jsou popsány metody aplikovány na rozsáhlejší soubory dat týkajících se studentů bakalářských oborů zapsaných v letech 2003/04 a 2004/05. Odstavec 2.2 je věnován odhadu průměrného prospěchu na VŠ. Alternativní výpočty vedou k závěru, že je pro tento odhad vhodné využít model se čtyřmi regresory (průměr známek ze střední školy, součet bodů za přijímací zkoušky, rok začátku VŠ studia a studijní program). S tímto výsledkem získaným aplikací lineárního modelu v podstatě souhlasí model získaný krokovou regresí. Další odstavce se zabývají predikcí úspěšnosti studia (úspěšné ukončení studia, úspěšné absolvování prvního ročníku, absolvování s vyznamenáním). Výsledky jsou většinou založeny na ROC analýze.

Zajímavé je porovnání výsledků dosažených autorkou s výsledky uváděnými v české i zahraniční literatuře. (Odst. 2.4)

Srovnání s českými pracemi: Závěry zpracované na fakultách, které jsou obdobné PřF UK, jsou podobné; výsledky Dr. Rubešové jsou v některých případech podrobnější a také přesnější. Autorka zmiňuje také práci týkající se predikční validity testu Obecných studijních předpokladů (OSP) společnosti Scio, s.r.o. Ukazuje se, že pro predikci je v tomto případě většinou významnější celkový výsledek testu OSP než jeho jednotlivé oddíly (verbální, analytický, kvantitativní). Vzhledem k tomu, že se rozšiřuje používání podobných testů připravovaných nejen společností Scio, bylo by asi účelné věnovat pozornost analýze souvislosti takových testů s možností predikce, už i proto, že testy tohoto typu začínají být součástí přijímacích zkoušek i na PřF UK.

Autorka porovnává své výsledky i se zahraničními studii, zejména z USA, kde má hodnocení přijímacích zkoušek delší tradici. Ukazuje se, že pokud jsou srovnatelné vstupní soubory, jsou srovnatelné i výsledky.

V závěrečné kapitole autorka stručně shrnuje výsledky, k nimž dospěla. Za nejdůležitější považují tvrzení obsažené v posledním odstavci, totiž to, že byla prokázána souvislost úspěšnosti studia na VŠ s předcházejícím prospěchem na střední škole. Na fakultách, kde počet uchazečů o studium určitého oboru nepřevyšuje výrazně kapacitu, by mohl dobrý prospěch na střední škole a úspěšné složení maturity stačit k přijetí ke studiu. Na PřF UK by mohl být tento postup aplikován v přijímacím řízení na některé obory, bylo by ovšem nutno pro prospěch na SŠ stanovit mez přijetí. Zkušenost ukazuje, že přijímání bez jakékoliv podmínky může způsobit i značné problémy.

Přílohy jsou věnovány popisu veličin v databázi student (A), simulaci (B), skriptům použitým pro výpočty v programu R (C). Texty skriptů jsou uvedeny na CD přiloženém k práci.

Práci hodnotím celkově jako významný příspěvek k problematice predikce úspěšnosti studia na vysoké škole, a to především na základě výsledků přijímacích zkoušek a prospěchu na střední škole. Je použita široká množina metod matematické statistiky, svědčící o dobré orientaci autorky v teorii těchto metod, přes to, že neabsolvovala vysokoškolské studium matematické statistiky. Rozsáhlé datové soubory pak jsou zpracovávány v různých kombinacích vstupních dat, výsledky jsou vědecky hodnoceny. Práce má velmi pěknou grafickou úroveň, obrázky i tabulky jsou pečlivě provedeny. Lze konstatovat, že předložená

práce je příkladem vědeckého řešení problému s využitím moderní teorie a že splňuje požadavky na práci disertační.

K práci mám pouze formální připomínky, které nikterak nesnižují její úroveň. Na str. 36 dole je zaměněno označení specificity a senzitivity. Problém „jednička“ kontra „jedna“. Podle mého mínění by se slangová vyjádření typu „doplněk do jedničky“, „funkce je rovna jedničce“ neměla objevovat ve vědeckém textu. Myslím, že mohla být věnována větší pozornost závěrečné korektuře textu, i proto, že se samozřejmě nelze v tak rozsáhlém a komplikovaném textu vyvarovat překlepů a nedopatření. Uvádím alespoň tyto:
V úvodním prohlášení autorky a v anglickém abstraktu – celkem čtyři nedopatření.
V seznamu obrázků je jiný popis než text pod obrázkem (např. obr. 1.1, 1.8, 2.2).
Někde se vloudila chybička při přepisu vzorců (např. ve vzorci 1.70)

Závěr. Předložená práce RNDr. Jany Rubešové je moderní vědecká práce věnovaná praktické užitečné problematice. Vzhledem k tomu, že splňuje předpoklady kladené na disertační práci, doporučuji, aby po její obhajobě byl autorce udělen titul doktor, PhD.

V Praze 29. srpna 2009