

Predikce úspěšnosti v testu matematických dovedností na základě rodinných faktorů

Teoretický úvod a cíl práce

Akademické úspěchy žáků a studentů ovlivňuje řada interpersonálních a intrapersonálních faktorů. Dominantní skupinou faktorů může být rodinné zázemí a socioekonomický status rodiny.

Z dosavadních výzkumů vyplývá, že mezi nejsilnější prediktory školního prospěchu patří především dosažené vzdělání rodičů, jejich očekávání ohledně dítěte a socioekonomický status rodiny. Vliv těchto proměnných potvrzuje řada studií např. Sánchez et al. (2013) ve své analýze identifikovali právě očekávání rodičů, úroveň jejich vzdělání, socioekonomický status a velikost rodiny jako hlavní prediktory školní úspěšnosti. Podobně Bullón et al. (2016) na vzorku 486 středoškoláků pozorovali, že akademické zázemí rodičů a jejich profesní zařazení významně predikují celkový výkon studentů.

Kromě objektivních demografických a statusových ukazatelů hrají zásadní roli také proměnné spojené s vnitřním fungováním rodiny. Midgett et al. (2004) ve svém výzkumu zahrnujícím 2 321 studentů zjistili, že vzdělání a očekávání rodičů mají přímý vztah ke známám, avšak působí i nepřímo prostřednictvím rodinné soudržnosti. Autoři rovněž zdokumentovali, že rodičovská vřelost, dohled a diskuse zaměřené na školu mají na prospěch prokazatelný vliv

Cílem předložené práce je prozkoumat, do jaké míry tyto faktory rodinného a domácího prostředí skutečně ovlivňují finální známku studentů na vzorku portugalských středoškolských studentů v matematice.

Zdroj dat a výzkumný soubor

Práce využívá volně dostupná data z datového repozitáře UCI Machine learning repository Kalifornské univerzity v Irvině, konkrétně data set Student Performance (Cortez, 2008).

Výzkumný soubor zahrnuje 396 portugalských studentek (52,5%) a studentů (47,5%) dvou středních škol ve věku od 15 do 22 let ($M=16,7$, $SD=1,2$), kterým na konci školního roku byla administrována dotazníková baterie (Cortez & Silva, 2008). Autoři původního výzkumu zjišťovali demografické údaje studentů, jejich rodin, skrz blíže nespecifikované dotazníkové metody posuzovali rodinou podporu, pomoc rodiny při vzdělávání, konzumaci návykových látek, rodinné vztahy, volnočasové aktivity studentů a výsledek zkoušky z matematiky.

Výsledkem zkoušky byla známka od 0 po 20, kdy vyšší hodnota znamená lepší výsledek. Průměrný výsledek zkoušení byl 10,41 se směrodatnou odchylkou 4,58.

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=673>.

Popis zkoumaných proměnných

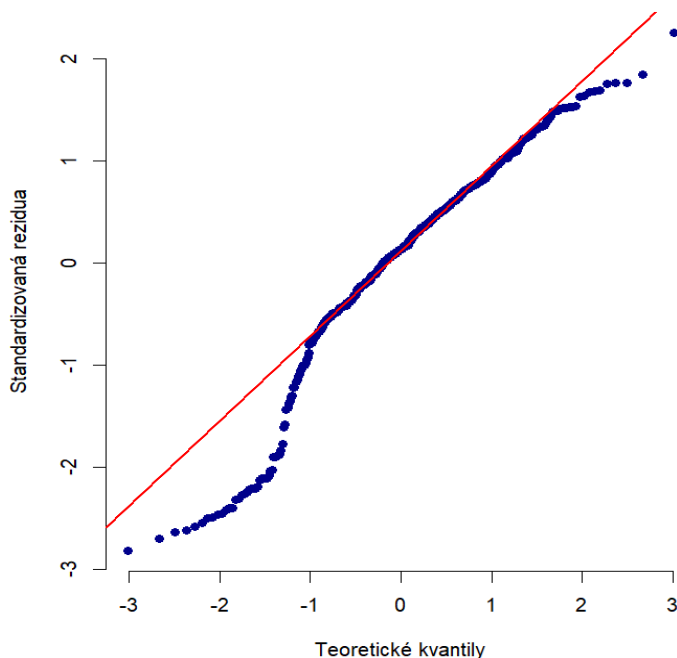
Závislá proměnná (konečná známka studenta, označovaná jako G3) nabývá v tomto datasetu hodnot od 0 do 20. Ačkoliv se jedná o diskrétní číselné hodnoty, pro účely této analýzy byla zvolena klasická mnohonásobná lineární regrese (metoda nejmenších čtverců – OLS).

Výzkumný soubor ($N = 395$) je dostatečně obsáhlý, aby naplňoval předpoklad centrální limitní věty, a tedy je vhodný pro výpočet lineární regrese. I přes drobná odchylení od normality reziduí u školních známek, která můžeme pozorovat níže na Grafu 1, je lineární model pro interpretaci vlivu rodinného zázemí stále vhodný

Nezávislé proměnné v tomto modelu jsou všechny proměnné z oblasti rodinného a domácího prostředí studenta. Patří mezi ně ordinální proměnné reprezentující vzdělání matky (Medu) a otce (Fedu), které byly kódovány od 0 po 4 (0=žádné vzdělání, 4= vysokoškolské vzdělání), kvalitu rodinných vztahů (famrel), která byla kódována od 1 po 5 (1= velmi špatné, 5= velmi dobré).

Další původně nominální a kategoriální proměnné, musely být pro účely regrese převedeny pomocí dummy kódování na sérii dichotomických proměnných (0 a 1). Typ bydliště (město/venkov), soužití rodičů (společné/oddělené), nadstandardní podpora žaka ze strany rodičů při vzdělávání (přítomna/nepřítomna), velikost rodiny (pod 3 osoby/nad 3 osoby), identita zákonného zástupce (matka/otec/ jiný) a profese obou rodičů rozdělené do několika kategorií (učitel, zdravotník, veřejný činitel, osoba v domácnosti, jiné).

Graf 1: *Q-Q graf reziduí lineárního modelu*



Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=673>.

Výsledky vícenásobné lineární regrese

Analýza dat byla provedena v programu RStudio verze 4.5.2 s využitím knihovny fastDummies pro dummy kódování kategoriálních proměnných.

Tabulka 1: Výsledky negativní binomické regrese

Prediktor	Odhad parametru	St. chyba	t-hodnota	p-hodnota
Počátek (Intercept)	7,01	1,86	3,76	<0.001
Vzdělání otce	0,10	0,29	0,33	0,74
Vzdělání matky	0,70	0,33	2,08	0,04
Rodinné vztahy	0,31	0,25	1,23	0,22
Bydliště: Město	0,67	0,56	1,21	0,228
Velikost rodiny: ≤ 3	0,74	0,52	1,43	0,15
Soužití rodičů: Spolu	-0,25	0,77	-0,32	0,75
Vzděl. rodinná podpora: Ano	-0,79	0,48	-1,65	0,10
Zástupce: Matka	-0,20	0,58	-0,35	0,73
Zástupce: Jiný	-1,21	0,94	-1,28	0,20
Práce matky: Zdravotník	1,40	1,14	1,223	0,22
Práce matky: Jiné	-0,02	0,74	-0,03	0,98
Práce matky: Veřejný činitel	0,82	0,82	1,00	0,32
Práce matky: Učitel	-0,11	1,05	-0,10	0,92
Práce otce: Zdravotník	0,46	1,50	0,31	0,76
Práce otce: Jiné	-0,11	1,07	-0,10	0,92
Práce otce: Veřejný činitel	-0,33	1,11	-0,30	0,77
Práce otce: Učitel	0,85	1,36	0,63	0,53

* Referenční kategorie pro jednotlivé dummy proměnné, vůči kterým jsou odhady parametrů vztahovány, byly stanoveny následovně: Bydliště: Venkov, Velikost rodiny: Více než 3 členové, Soužití rodičů: Odděleně, Rodinná podpora při studiu: Ne, Zákonný zástupce: Otec, Profese matky: V domácnosti, Profese otce: V domácnosti

Na základě výsledků lineární regrese prezentovaných v Tabulce 1 můžeme pozorovat, že naprostá většina proměnných rodinného zázemí nepředstavuje statisticky významný prediktor pro finální známku studenta (p -hodnoty $> 0,05$).

Vliv bydliště, kvality vztahů, velikosti rodiny ani profese rodičů se v tomto data setu nepotvrdil. Jediným statisticky významným faktorem je vzdělání matky ($p = 0,038$). Odhad parametru ukazuje, že s každým dosaženým stupněm vzdělání matky se finální známka studenta zvyšuje průměrně o 0,695 bodu.

Zajímavým zjištěním je hraniční statistická významnost u rodinné podpory při studiu ($p = 0,09$), která ovšem vykazuje negativní vliv (odhad -0,79). To naznačuje, že je hypoteticky možné, že studenti s extra podporou ze strany rodiny dosahují průměrně horších výsledků, což lze

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=673>.

pravděpodobně vysvětlit tím, že tato podpora je častěji poskytována studentům, kteří již mají akademické potíže a nadaní studenti podporu navíc nepotřebují a možná nemají, takže nemá na první pohled očekávatelný pozitivní efekt.

Kvalita modelu

Kvalita navrženého modelu byla ověřena pomocí koeficientu determinace (R^2) a testové statistiky F. Model jako celek se vykazuje jako statisticky významný (F-statistika = 2,141, $p = 0,005466$), což znamená, že vybrané nezávislé proměnné mají prokazatelný vztah k závislé proměnné.

Hodnota koeficientu determinace (R^2) však dosahuje pouze 0,08804. Zjištěná hodnota plně koresponduje s výstupy regrese z Tabulky 1 a potvrzuje, že samotné rodinné zázemí (vzdělání a zaměstnání rodičů, bydliště, kvalita rodinných vztahů) dokáže vysvětlit necelých 9 % celkové variability známky z matematiky jednotlivých studentů. Většina rozdílů v hodnocení studentů je tedy dána jinými faktory mimo rodinné zázemí.

Závěr

Jediným prokazatelným pozitivním prediktorem úspěchu v matematice na souboru portugalských studentů SŠ se ukázalo být dosažené vzdělání matky. Naopak přítomnost extra rodinné podpory při vzdělávání vykazovala hraničně významný negativní vliv, což je možné vysvětlit jevem, kdy je dodatečná pomoc ze strany rodiny nabízena až ve chvíli, kdy student začne akademicky selhávat.

Navržený regresní model vysvětlil pouze necelých 9 % ($R^2 = 0,088$) variability finální známky z matematiky. Z toho vyplývá, že ačkoliv rodinné zázemí poskytuje studentům možný odrazový můstek, většina rozdílů ve studijních výsledcích je formována jinými vlivy.

Seznam literatury

Cortez, P. (2008). *Student Performance*. UCI Machine Learning Repository. <https://doi.org/10.24432/c5tg7t>

Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. *EUROSIS*.

Fajardo Bullón, F., Maestre Campos, M., Felipe Castaño, E., León Del Barco, B., & Polo Del Río, M. I. (2016). ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LOS ALUMNOS DE EDUCACIÓN SECUNDARIA OBLIGATORIA SEGÚN LAS VARIABLES FAMILIARES. *Educación XXI*, 20(1). <https://doi.org/10.5944/educxx1.17509>

Midgett, J., Ryan, B. A., & Adams, G. (2004). *Family Processes, Individual Characteristics, and Adolescent Academic Achievement: A Study of the Ecology of Family-School Connections*. <https://www.semanticscholar.org/paper/Family-Processes%2C-Individual-Characteristics%2C-and-A-Midgett-Ryan/81917cad6c993d519634d0798eea66321c80cd2f>

Pérez Sánchez, C. N., Betancort Montesinos, M., & Cabrera Rodríguez, L. (2013). Influencias de la familia en el rendimiento académico. Un estudio en Canarias. *Revista Internacional de Sociología*, 71(1), 169–187. <https://doi.org/10.3989/ris.2011.04.11>