

ANALÝZA VLIVU PROMĚNNÝCH NA PŘIJETÍ NA VYSOKOU ŠKOLU U AMERICKÉHO STUDENTA

V tomto úkolu se budu věnovat analýze vlivu proměnných na přijetí na vysokou školu u amerického studenta. K dispozici máme data 400 respondentů, od každého respondenta jsme získali tyto údaje:

- Zdali byl či nebyl přijat na vysokou školu
- GRE (Graduate Record Exam scores) - standardizovaný test, který slouží k hodnocení schopností uchazečů o studium
- GPA - průměr známek studenta vyjádřený na stupnici od 1 do 4, kde známka A = 4,0. Tento ukazatel se používá často jako kritérium pro přijetí na vysokou školu, kdy vyšší GPA může vést k většímu počtu nabídek na přijetí
- Prestiž středoškolské instituce, renomé a kvality vzdělávací instituce, kde 1 značí nejvyšší prestiž.

Tato data byla získána z volně dostupných datasetů. V tabulce uvádím základní deskriptivní statistiky ukazatelů.

	GRE	GPA	Prestiž
Průměr	587,700	3,351	2,485
Medián	580	3,365	2
Min	220	2,260	1
Max	800	4	4

Pokusím se tedy modelovat vztah proměnných **GRE**, **GPA** a **prestiž** (nezávisle proměnné) na **přijetí** či **nepřijetí studenta na VŠ** (závislá proměnná) pomocí **logistické regrese**.

Data analyzuji v programu TIBCO Statistica, verze 14.

V tabulce níže vidíme odhadnuté parametry, p-hodnoty a hodnoty Waldova testu. Všechny parametry můžeme označit za **významné**, můžeme zamítnout nulovou hypotézu o tom, že by parametr byl rovný nule.

	Odhadnutý parametr	Waldova statistika	p-hodnota
(počátek)	-3,450	9,272	0,002
GRE	0,002	4,414	0,036
GPA	0,777	5,630	0,018
Prestiž	-0,560	19,404	0,000

V další tabulce uvádím poměry šancí.

Poměr šancí

(počátek)	
GRE	1,002
GPA	2,175
Prestiž	0,571

Jak vidíme, hodnota u GRE je velmi blízká nule, tedy můžeme říct, že GRE hraje velmi malou, až nulovou roli v přijetí na VŠ. Hodnota u GPA nám říká, že se šance zdvojnásobuje s každým bodem GPA navíc. Hodnota u proměnné *prestiž* nám říká, že je s každým zvýšením prestiže o jednotku poloviční šance na přijetí. Tady si ale musíme uvědomit, že „zvýšení prestiže“ jako proměnné o jednotku znamená, že škola má prestiž nižší. Správně tedy musíme interpretovat tuto hodnotu takto: pokud dojde k poklesu prestiže o 1 bod, je šance na přijetí o polovinu nižší. Studentům bychom tedy doporučili se zaměřit především na zvýšení GPA během studia.

Pokud bychom hodnotili kvalitu modelu jako celku, nejprve se podíváme na p-hodnotu Waldova testu pro celý model. Vidíme, že $< 0,0001$, tedy zamítáme nulovou hypotézu, že prognózy modelu se liší od náhody. Pro úplnost ještě uvedeme hodnotu AIC, která je 467,442 a hodnota BIC je 483,407. Koeficient determinace, respektive jeho analogie, Nagelkerke R² je roven 0,135. Pro úplnost uvádím tabulku srovnání počtu přijatých a nepřijatých versus předpovědi dle modelu

		Předpověď		Procento správných
		Přijat	Nepřijat	
Pozorování	Přijat	29	98	22,835
	Nepřijat	20	253	92,674

Na závěr můžeme ještě zvážit, zdali by se tento model nedal upravit tak, že proměnnou *prestiž* bychom klasifikovali jako kategorickou. Toto si ale necháme pro další výzkumníky na prozkoumání.

Zdroje

UCLA: IDRE (Institute for Digital Research and Education). Data Analysis Examples. from <http://www.ats.ucla.edu/stat/dae/>