

Logistická regrese

Predikovaná proměnná: pohlaví

Prediktory:

- Důvěrnost s rodiči
- Důvěrnost s přáteli
- Životní spokojenost
- Očekávané vzdělání
- Oblíbený předmět – kódován jako kategorická proměnná

Vzorek: 703 dětí staršího školního věku a adolescentů. Původní vzorek byl 768 dětí staršího školního věku a adolescentů, takže došlo k docela výrazné redukcí původního vzorku. U 65 jedinců chyběla jedna nebo více hodnot prediktorů, u dvou dokonce hodnota predikované proměnné. Abychom získaly co nejpřesnější výsledky, těchto 65 jedinců jsme z regrese vyloučily.

Jako predikovanou proměnnou jsme zvolily dichotomickou kategorickou proměnnou pohlaví s hodnotami muž a žena. Jako prediktory jsme zvolily proměnné, o nichž se domníváme, že pro obě pohlaví nabývají různých hodnot a proto by mohly mít na určení pohlaví určitý podíl.

Tabulka 1. Popisné statistiky proměnných a korelace.

Prediktor	M	SD	Min	Max	Korelace (Pearson)			
					Oblíbený předmět	Očekávané vzdělání	Důvěrnost s rodiči	Důvěrnost s přáteli
Oblíbený předmět	19,28	24,36	1	98				
Očekávané vzdělání	4,98	1,35	1	6	-,014			
Důvěrnost s rodiči	2,84	,56	1,33	4,00	-,085*	,095**		
Důvěrnost s přáteli	2,70	,51	1,17	4,00	,036	,032	-,054	
Životní spokojenost	2,91	,48	1,00	4,00	-,031	,218**	,401**	,012

** Korelace je signifikantní na 0,01 hladině významnosti (oboustranně).

* Korelace je signifikantní na 0,05 hladině významnosti (oboustranně).

Proměnné důvěrnost s rodiči, důvěrnost s přáteli a životní spokojenost jsou spojitě proměnné a mohou nabývat hodnot 1 až 4. Proměnná očekávané vzdělání nabývá hodnot 1 až 6, přičemž 1 je základní vzdělání (zde nejnižší možné) a 6 je vysoká škola (zde nejvyšší možné).

Proměnnou oblíbený předmět jsme kódovaly jako kategorickou proměnnou. Předmět český jazyk jsme při kódování určily jako základní skupinu.

Analýza

Logistická regresní analýza byla provedena mezi dichotomickou závislou proměnnou (pohlaví) a nezávislými proměnnými (důvěrnost s rodiči, důvěrnost s přáteli, životní spokojenost, očekávané vzdělání, oblíbený předmět - kódován jako kategorická proměnná). Analýza byla provedena za použití programu SPSS Statistics 21.

Proměnné měřené alespoň na intervalové škále byly u obou hodnot závislé proměnné přibližně normálně rozložené a homoskedascita reziduí byla dodržena.

Logistická regresní analýza odhalila, že model statisticky významně předpovídal pohlaví respondentů, $R^2 = 0,22$ (Cox & Snell); 0,30 (Nagelkerke); $\chi^2 (26, N = 703) = 174,56$; $p < 0,001$. Tabulka 2 zobrazuje nestandardizované regresní koeficienty (B), jejich směrodatné chyby a signifikanci, Exp (B) a 95% interval spolehlivosti pro Exp(B) pro každou proměnnou.

Tabulka 2. Výsledky logistické regresní analýzy.

Prediktor	B	SE	Wald		Exp(B)	95% CI pro Exp(B)	
			χ^2	p		Min	Max
Očekávané vzdělání	0,16	0,07	5,332	,038	1,18	1,03	1,35
Důvěrnost s rodiči	0,37	0,17	4,562	,038	1,45	1,03	2,04
Důvěrnost s přáteli	1,35	0,19	50,526	,038	3,85	2,65	5,58
Oblíbený předmět							
Životní spokojenost	-0,58	0,21	7,350	,038	0,56	0,37	0,85
Anglický jazyk	-0,39	0,75	,267	,606	0,68	0,16	2,95
Německý jazyk	-0,12	0,86	,019	,890	0,89	0,16	4,82
Matematika	-1,15	0,68	2,831	,092	0,32	0,08	1,21
Fyzika	-2,37	0,93	6,470	,011	0,09	0,02	0,58
Chemie	-2,62	0,90	8,494	,004	0,07	0,01	0,42
Biologie	-0,92	0,74	1,526	,217	0,40	0,09	1,71
Zeměpis	-0,83	0,78	1,129	,288	0,44	0,09	2,01
		27586,1					
Základy soc. věd	19,14	0	,000	,999	205286058,62	0,00	
Dějepis	-1,32	0,69	3,732	,053	0,27	0,07	1,02

Pracovní vyučování	-1,53	1,13	1,824	,177	0,22	0,02	1,99
Tělesná výchova	-1,87	0,65	8,282	,004	0,15	0,04	0,55
Výtvarná výchova	1,12	0,83	1,802	,179	3,06	0,60	15,70
Hudební výchova	-0,67	0,83	,657	,418	0,51	0,10	2,60
Rodinná výchova	-0,92	0,85	1,165	,280	0,40	0,08	2,11
Chemické laboratoře	19,80	5	,000	,999	397208654,82	0,00	
Technické kreslení	19,67	7	,000	1,000	348607660,23	0,00	
Občanská nauka	0,37	1,29	,083	,773	1,45	0,12	18,06
Ekonomika	-0,88	1,21	,527	,468	0,42	0,04	4,46
Informatika	-3,33	1,26	6,940	,008	0,04	0,00	0,43
Ekologie	18,92	6	,000	,999	164283780,25	0,00	
Deskriptivní geometrie	-	40192,9					
Jiné	22,30	7	,000	1,000	0,00	0,00	
Jiné	-1,39	0,69	4,042	,044	0,25	0,06	0,97

Vysoké hodnoty Exp (B) u některých proměnných souvisí s vysokými hodnotami směrodatné chyby.

Model dokázal vysvětlit **71% sdíleného rozptylu**. Ačkoliv se nám nepodařilo vysvětlit větší část rozptylu (bez prediktorů bylo vysvětleno 59%), zjistily jsme, že u adolescentů z našeho vzorku se na vysvětlení predikované proměnné, tedy pohlaví, podílí nejvíce zejména důvěrnost s přáteli a překvapivě také životní spokojenost.

Zdroj dat:

Ke splnění tohoto úkolu jsme využily data Long1, umístěná ve studijních materiálech předmětu Statistická analýza dat II (<https://is.muni.cz/auth/el/1423/podzim2013/PSY252/um/>).

V té interpretaci jste se mohly rozepsat víc. Taky předpoklady zůstaly nezmněny.