

Karolína Popelková 397556

Jiří Sláma 274356

Úkol č. 3: Logistická regrese

Pro účely tohoto úkolu, ve kterém jsme měli analyzovat data pomocí logistické regrese, jsme si vybrali datový soubor *EU Kids Online*. Jedná se o výzkum, jehož autory jsou Livingstone S., Haddon L., Görzig A. a Ólafsson, K. z roku 2011. Výzkumníci se zde zaměřili na to, jak mladí lidé (zejména děti) užívají nová média. Konkrétně s jakými ohroženími se děti na internetu setkávají, jaké je podvědomí jejich rodičů o těchto potenciálních problémech, jaký mají přehled o svém dítěti, co dělá, když je na internetu a identifikovali jednotlivé typy riskantního chování na internetu. Jedná se o velmi rozsáhlý výzkum napříč 25 zeměmi, celkem se ho zúčastnilo 25. 142 dětí a pokaždé jeden z jejich rodičů.

My jsme se zaměřili na tyto prediktory:

- *Dítě v uplynulém roce na internetu shlédlo agresivní nebo násilné obrázky zobrazující útok nebo zabití druhé osoby.* Odpovědi ano/ne. (N = 16 053) (pozn.: Agresivní obrázky)
- *Dítě v uplynulém roce shlédlo na internetu obrázky, které byly zcela jasně sexuálního charakteru. Například zobrazení nahých osob, nebo osob při sexu.* Odpovědi ano/ne. (N = 14 914) (pozn.: Sexuální obrázky)
- *Když přemýšlíte o svém dítěti, kterých z těchto věcí se obáváte nejvíce? Že se dítě dostane do problémů s policií.* Odpovědi ano/ne. (N = 18 706) (Problémy s policií)
- *Jak moc pravdivé je pro tebe tvrzení: dělám nebezpečné věci kvůli zábavě.* Odpovědi nepravdivé/trochu pravdivé/velmi pravdivé. (N = 18 577) (pozn.: Nebezpečí)
- *V posledním měsíci jsem hrál PC hry proti sobě nebo proti PC.* Odpovědi ano/ne. (N = 15 371). (pozn.: PC hry)

Naším záměrem je z daných prediktorů předpovědět pravděpodobnost, že respondent je muž nebo žena.

Daná data splnila veškeré předpoklady pro provedení logistické regrese (linearita, nezávislost odchylek, nepřítomnost multikolinearity). Z výsledků vyplývá, že daný model je signifikantní, $\chi^2(1, N = 18 706) = 76,46, p < 0, 001$. Jednotlivé prediktory jsou ovšem schopné určit pravděpodobnost pohlaví pouze z přibližně 10 % (podle Nagelkerke R^2).¹ Navíc *Odd Ratio*

¹ Jsme si vědomi, že tyto statistiky jsou pouze přibližným odhadem, avšak poskytují alespoň rámcový obraz o zvoleném modelu.

Karolína Popelková 397556

Jiří Sláma 274356

Úkol č. 3: Logistická regrese

jednotlivých prediktorů indikuje, že šance na správné určení závislé proměnné (že jev nastane) je ještě nižší, jelikož poměry jsou ve většině případů menší než 1 (viz. Tabulka 1).

Slabý vztah prediktorů k závislé proměnné by mohl být způsoben velkým vzorkem respondentů, v takovém případě lze předpokládat, že rozdíly mezi pohlavími budou minimální, bez ohledu na proměnnou. Je nutné ovšem vzít v úvahu, že pokud předpokládáme, že tento model predikoval pohlaví z přibližně 10%, tak je to vzhledem ke všem možným existujícím proměnným, ze kterých by se dalo usuzovat na pohlaví, poměrně vysoké číslo. Navíc testovaný model vylepšuje správný odhad pohlaví z původních 52,5 % na 61,7 %. Pokud se zaměříme na rozdíl predikujícího modelu oproti původnímu pomocí chi-kvadrátu, model s prediktory snížil -2LL (14750,7) o 905,24. Svůj vliv na lepší predikci pohlaví tedy rozhodně má. Výsledky statistiky *Odd Ratio* ukazují, že nevyšší šance na určení pohlaví v tomto modelu je z proměnných, ve kterých rodiče odpovídají na otázky, jestli jejich děti v minulém roce viděli na internetu obrázky s agresivním nebo sexuálním obsahem.

Tabulka 1: Logistická regrese předpovídající pohlaví z proměnných Agresivní obrázky, Sexuální obrázky, Problémy s policií, Nebezpečí a PC hry

Prediktor	β	SE(β)	Wald χ^2	p	Odds Ratio	Interval Spolehlivosti	
Agresivní obrázky	0,49	0,05	107,41	< 0,001	1,64	1,49	1,80
Sexuální obrázky	0,23	0,05	17,95	< 0,001	1,25	1,13	1,39
Problémy s policií	-0,34	0,06	37,58	< 0,001	0,71	0,64	0,79
Pc hry	-0,93	0,05	308,71	< 0,001	0,39	0,36	0,44
Nebezpečí							
Trochu pravdivé	-0,62	0,05	158,52	< 0,001	0,54	0,49	0,60
Velmi pravdivé	-0,90	0,11	70,33	< 0,001	0,41	0,33	0,50

Tabulka 2: Relativní četnosti prediktorů vzhledem k pohlaví

	Pohlaví	
	Muži	Ženy
Agresivní obrázky	50%	50%
Sexuální obrázky	49%	51%
Problémy s policií	50%	50%
Pc hry	53%	47%
Nebezpečí	50%	50%

Karolína Popelková 397556

Jiří Sláma 274356

Úkol č. 3: Logistická regrese

Zdroje

Livingstone, S., Haddon, L., Görzig, A., and Ólafsson, K. (2011). *Risks and safety on the internet: The perspective of European children. Full Findings*. LSE, London: EU Kids Online.