

Seminární úkol č. 3

Autoři: Klára Čapková (406803), Markéta Peschková (414906)

Zdroj dat: EU Kids Online Survey

Popis dat

Analyzovaná data pocházejí z výzkumu online chování dětí Evropské unie. Data shrnují údaje o 18709 respondentech ve věku 11 – 16 let, z nichž 9352 bylo mužského a 9357 ženského pohlaví. Analýza z důvodu chybějících dat u některých položek pracuje pouze s 16003 případy.

Komentář [SJ1]: děti matičky Rusi...

Popis modelu

Logistická regrese byla provedena za účelem predikování pravděpodobnosti toho, že sledovaný jedinec v posledním měsíci hrál online hry přes internet s jinými lidmi. Tato proměnná byla predikována pomocí sedmi nezávislých proměnných.

Komentář [SJ2]: Její četnosti bychom taky chtěli znát.

Komentář [SJ3]: raději prediktorů

První z proměnných, které jsme do analýzy zahrnuli, je pohlaví dítěte. Předpokládáme, že u chlapců je větší pravděpodobnost hraní online počítačových her než u dívek. Chlapci a dívky byli v naší analýze zastoupeni téměř zcela rovnoměrně, chlapců bylo 8006 a dívek 7997.

Komentář [SJ4]: Pro velká N jsou asi lepší %.

Další proměnná „Different online persona“ udává, do jaké míry si jedinec vytváří na internetu novou identitu, odlišnou od té, která je pro něj typická ve skutečném světě. Jedinci, kteří spíše inklinují k vytváření nové identity, budou mít dle našeho názoru vyšší schopnost imaginace, a tudíž budou rádi trávit čas hraním online her, kde je hráči často přiřazena určitá role, kterou má naplnit, a to navíc v nereálném světě.

Dále jsme do modelu zahrnuli i proměnnou „online aktivity povoleny kdykoli“, u níž předpokládáme, že rodičovské svolení k neomezenému užívání internetu bude dítěti umožňovat častější hraní online her.

U proměnné „celkový počet online aktivit“, která udává, jak moc se sledovaný jedinec věnuje činnostem na internetu, předpokládáme, že hraní počítačových online her bude zaznamenáno spíše u jedinců s vyšším počtem online aktivit. Popisné statistiky těchto 3 spojených proměnných zobrazuje Tabulka 1.

Proměnnou „stahování hudby a filmu“, která udává, jestli si dítě v posledním měsíci z internetu stáhlo hudební nebo filmový materiál, jsme mezi prediktory zařadily, jelikož předpokládáme, že vyšší frekvence stahování by opět mohla ukazovat na častější hraní online her.

Stejně tomu je i u proměnné „trávení času ve virtuálním světě“, která podává informaci o tom, zda se jedinec v posledním měsíci této aktivitě věnoval.

To, zda dítě v posledním roce navštívilo chatovací místnost, by také mohlo být dobrým prediktorem hraní her na internetu. Četnosti těchto 3 kategoričkových proměnných zobrazuje Tabulka 2 a Graf 1.

Tabulka 1

Popisné statistiky spojitych proměnných Online aktivity povoleny kdykoli, Different online persona a Celkový počet online aktivit

Proměnná	M	SD	Min	Max
Online aktivity povoleny kdykoli	3,36	3,31	-2	12
Different online persona	1,59	1,56	0	6
Celkový počet online aktivit	8,13	3,47	0	17

Tabulka 2

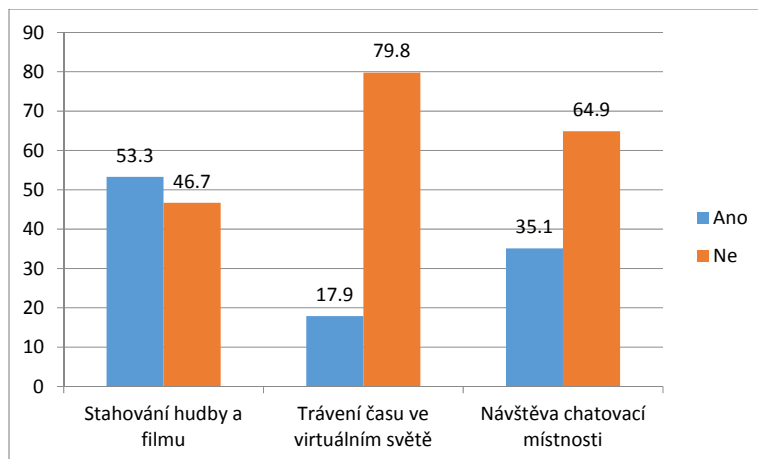
Tabulka četností pro kategorické proměnné Stahování hudby a filmu, Trávení času ve virtuálním světě a Návštěva chatovací místnosti

Proměnná	Procenta	
	Ano	Ne
Stahování hudby a filmu	53,30	46,70
Trávení času ve virtuálním světě	17,9	79,80
Návštěva chatovací místnosti	35,10	64,90

Komentář [SJ5]: Obvykle stačí jen procenta v textu.

Graf 1

Graf četností (%) pro proměnné Stahování hudby a filmu, Trávení času ve virtuálním světě a Návštěva chatovací místnosti



Komentář [SJ6]: Když ta procenta dávají 100%, jsou lepší stacked sloupce. Pak by tam šlo zobrazit i rozdělení podle závislé.

Logistická regrese

Wald test na $\alpha=0,05$ byl pro všechny nezávislé proměnné signifikantní, můžeme tedy říci, že každý z prediktorů přispívá ke zlepšení modelu. Nejvýznamnějším prediktorem modelu je proměnná stahování hudby a filmů, která sama o sobě napomáhá ke správnému klasifikování 9,1%. Dalšími

Komentář [SJ7]: Teď už vím, že jste to po jednotlivých prediktorech krokovaly. Pak je to pořadí vkládání třeba popsat a zdůvodnit.

dobrymi prediktory jsou promenne *pohlavi* a *celkovy pocet aktivit online*. Kazda z nich zlepšuje kvalitu modelu o cca 5 %. Nejmeňe přispivaji zlepšení modelu promenne *Different Online Persona* a *navštěva chatovací místnosti*.

Test celkového modelu oproti modelu pouze s průměrem byl signifikantní, $\chi^2 = (6, 16003) = 5534,06$, $p < 0,05$. Nagelkerke $R^2 = 0,39$, Cox & Snell $R^2 = 0,29$. Model správně klasifikoval 75,6 % těch, kteří v minulém měsíci nehráli na internetu online hry a 72,5 % jedinců, kteří v minulém měsíci hráli na internetu online hry, tj. celkově byl model úspěšný v 74,1 %. Blok 0, který klasifikoval jedince před zahrnutím vybraných prediktorů do analýzy, tedy pouze pomocí konstanty, klasifikoval správně 50,3 % jedinců. Celkově náš model logistické regrese tedy zvýšil efektivitu predikce o 23,8 %.

Výsledky logistické regrese pro jednotlivé prediktory shrnuje Tabulka 3.

Tabulka 3

Logistická regrese predikující hraní online her přes internet ze zvolených nezávislých proměnných

Prediktor	β	SE	Exp(B)	Wald test
Konstanta	-2,78	0,08	0,06	1132,87
Pohlaví dítěte	-1,39	0,04	0,25	1267,04
Stahování hudby a filmu	-0,36	0,04	0,70	65,58
Trávení času ve virtuálním světě	0,56	0,06	1,75	99,40
Online aktivity povoleny kdykoli	0,06	0,01	1,06	66,60
Index Different Online Persona	0,04	0,01	1,04	8,31
Návštěva chatovací místnosti	-0,21	0,04	0,82	21,31
Celkový počet online aktivit	0,41	0,01	1,50	1797,34

Komentář [SJ8]: Tohle je nestandardizovaný koeficient – B.

Komentář [SJ9]: Jaký význam má tohle číslo?

Komentář [SJ10]: Oproti interpretaci níže se mi nezdá, že by tohle byl malý efekt – téměř dvojnásobná šance.

Komentář [SJ11]: Když Wald, tak aspoň s hvězdičkama nebo s p.

Níže naleznete rovnici logistické regrese s hodnotami prototypického jedince, tj. jedince, který vykazuje průměrné hodnoty spojitých proměnných a častější hodnoty dichotomických proměnných. Takovým prototypickým jedincem je tedy chlapec, který stahuje filmy a hudbu, netráví čas ve virtuálním světě, nenavštěvuje chatovací místnost, dosahuje skóru 3,36 na škále zobrazující, do jaké míry mu jsou povoleny online aktivity kdykoli, dále skóru 1,60 na škále Different online osoby a oddává se celkem cca 8 online aktivitám. Pravděpodobnost, že takový jedinec hrál v minulém měsíci online hry je 0,61.

Komentář [SJ12]: Účel počítání těch pravděpodobností jsme si vyjasnili na semináři..

$$\ln O_{y=\text{hraní online her}} = -2,78 - 1,39 \text{ pohlaví} - 0,36 \text{ stahování filmů a hudby} + 0,56 \text{ čas ve virtuálním světě} + 0,06 \text{ online aktivity kdykoli} + 0,04 \text{ Online different persona} - 0,21 \text{ chatovací místnost} + 0,41 \text{ celkem online aktivit}$$

$$\ln O_{y=\text{hraní online her}} = -2,78 - 1,39 * 0 - 0,36 * 1 + 0,56 * 0 + 0,06 * 3,36 + 0,04 * 1,60 - 0,21 * 0 + 0,41 * 8,13 = 0,46$$

$$P_{Y=\text{hraní online her}} = \frac{1}{1 + e^{-0,46}} = 0,61$$

Interpretace výsledků

U proměnných Trávení času ve virtuálním světě, Online aktivity povoleny kdykoli, Index Different Online Persona a Celkový počet online aktivit se tedy naplnily naše předpoklady, že vyšší skóry (případně kladné odpovědi) v těchto proměnných zvyšují pravděpodobnost, že jedinec hraje online hry, avšak toto zvýšení nebylo tak výrazné, jak jsme očekávaly.

Naším předpokladům odpovídá také zjištění, že je-li jedinec mužského pohlaví, je šance, že hraje online hry čtyřnásobná oproti případu, kdy se jedná o dívku.

U následujících dvou proměnných se však naše předpoklady nepotvrdily. V případě, kdy respondent uvedl, že v posledních dvanácti měsících navštívil chatovací místnost, šance na to, že je online hráčem, jsou 1,22krát menší, což je pro nás poměrně překvapivé.

Stejně je tomu i u proměnné Stahování hudby a filmu, jejíž kladná hodnota snižuje šance na to, že jedinec hraje hry přes internet, a to 1,43krát.

Tuto skutečnost lze vysvětlit například tak, že dítě, které navštěvuje chatovací místnosti, může být spíše orientované na sociální komunikaci s jinými osobami. Hraní online her by možná nenaplnovalo dostatečně jeho potřebu sociální interakce, protože při hře je komunikace s ostatními hráči upozaděna, není hlavní náplní hry. V případě stahování hudby a filmů si můžeme překvapivé výsledky vysvětlit tak, že jedinec, který tato multimédia stahuje, pravděpodobně vyhledává takovou náplň volného času, při níž není nutné navazovat kontakt s jiným jedincem. Je také možné, že pro takového jedince je internet pouze prostředkem k získání těchto souborů.

Závěrem je třeba kriticky pohlédnout na námi navržený model a zhodnotit jeho kvalitu. Vezmeme-li v úvahu, že jsme do modelu zahrnuly vedle pohlaví pouze proměnné, které přímo souvisí s užíváním internetu (s typem a rozsahem činností, které jedinec na internetu vykonává), můžeme nárůst správně klasifikovaných jedinců o 23,8 % považovat za uspokojivý. Ke správnému klasifikování zbylých 26 % jedinců, které náš model nepokrývá, by pravděpodobně mohly přispět nezávislé proměnné, které by se soustředily na osobnostní rysy jedince, jeho zaměření apod.

Dobrý první pokus, včetně úvodního zdůvodnění a závěrečné interpretace. Chybí jen vztahy se závislou v deskriptivách a nějaká zmínka o předpokladech.

Komentář [SJ13]: +!

Komentář [SJ14]: Bravo, dobře převrátily!