

Seminární práce č. 3: Logistická regrese

V této seminární práci jsme vycházeli z EUKO dat. Pokoušeli jsme se predikovat, zdali dítě má nebo nemá profil na sociální síti, čili odhadnout hodnotu proměnné „Máš v současnosti profil na sociální síti, který používáš, nebo ne?“ na základě nezávislých proměnných:

- „Jak často používáš internet?“
- „Myslíš, že jsou na internetu věci, které jedince tvého věku obtěžují?“
- „Kolik myslíš, že toho rodiče ví o tom, co děláš na internetu?“
- „Ignoroval jsi někdy, co ti rodiče řekli, když jsi používal internet, nebo ne?“
- „Vím více o internetu, než má rodiče“
- „Vím hodně o používání internetu“

a

- „Myslíš, že je na internetu hodně dobrých věcí pro děti tvého věku?“

Vzhledem ke skutečnosti, že všechny naše proměnné byly kategorické, neověřovali jsme předpoklady linearity, **multikolinearity a nezávislosti chyb**.

Následují tabulky popisných statistik každé proměnné.

Tabulky deskriptivních statistik

Tabulka 1.1

Deskriptivní statistiky kategorické proměnné „Máš v současnosti profil na sociální síti, který používáš, nebo ne?“

Máš v současnosti profil na sociální síti, který používáš, nebo ne?

	Četnosti	Relativní četnosti
Ano	13518	72,3
Ne	5114	27,3
Celkem platné	18632	99,6
Nevím (chybějící hodnoty)	77	0,4
Celkem	18709	

Tabulka 1.2

Deskriptivní statistiky kategorické proměnné „Jak často používáš internet?“

Jak často používáš internet?

	Četnosti	Relativní četnosti
Každý den nebo téměř každý den	13730	73,4
Jednou či dvakrát za týden	4258	22,8
Jednou či dvakrát za měsíc	497	2,7
Méně než jednou za měsíc	159	0,8
Celkem platné	18644	99,7
Nevím (chybějící hodnoty)	65	0,3
Celkem	18709	

Tabulka 1.3

Deskriptivní statistiky kategorické proměnné „Vím více o internetu, než mí rodiče“

Vím více o internetu, než mí rodiče

	Četnosti	Relativní četnosti
Není pravda	4100	21,9
Částečně pravda	5921	31,6
Zcela pravda	8326	44,5
Celkem platné	18347	98,1
Chybějící	362	1,9
Celkem	18709	

Tabulka 1.4

Deskriptivní statistiky kategorické proměnné „Víš toho hodně o používání internetu?“

Vím hodně o používání internetu

	Četnosti	Relativní četnosti
Není pravda	1951	10,4
Částečně pravda	9029	48,3
Zcela pravda	7539	40,3
Celkem platné	18519	99,0
Chybějící	190	1,0
Celkem	18709	

Tabulka 1.5

Deskriptivní statistiky kategorické proměnné „Myslíš, že je na internetu hodně dobrých věcí pro děti tvého věku?“

Na internetu je mnoho věcí, prospěšných pro děti mého věku

	Četnosti	Relativní četnosti
Není pravda	1187	6,3
Částečně pravda	7972	42,6
Zcela pravda	8956	47,9
Celkem platné	18115	96,8
Chybějící	594	3,2
Celkem	18709	

Tabulka 1.6

Deskriptivní statistiky kategorické proměnné „Myslíš, že jsou na internetu věci, které jedince tvého věku obtěžují?“

Myslíš, že jsou na internetu věci, které jedince tvého věku obtěžují?

	Četnosti	Relativní četnosti
Ano	11100	59,3
Ne	6367	34,0
Celkem platné	17467	93,4

Nevím (chybějící hodnoty)	1242	6,6
Celkem	18709	

Tabulka 1.7

Deskriptivní statistiky kategorické proměnné „Kolik myslíš, že toho rodiče ví o tom, co děláš na internetu?“

Kolik myslíš, že toho rodiče ví o tom, co děláš na internetu?

	Četnosti	Relativní četnosti
Hodně	4992	26,7
Celkem dost	7268	38,8
Pouze trochu	4638	24,8
Nic	1235	6,6
Celkem platné	18133	96,9
Chybějící hodnoty	1	0,0
Nevím	575	3,1
Celkem	18134	

Tabulka 1.8

Deskriptivní statistiky kategorické proměnné „Doporučoval ti někdy některý z tvých rodičů, jak používat internet bezpečně?“

Doporučoval ti někdy některý z tvých rodičů, jak používat internet bezpečně?

	Četnosti	Relativní četnosti
Ano	7191	38,4
Ne	10690	57,1
Celkem platné	17881	95,6
Nevím	827	4,4
Chybějící hodnoty	1	0,0
Celkem	18709	

Tabulka 1.9

Deskriptivní statistiky kategorické proměnné „Ignoroval jsi někdy, co ti rodiče řekli, když jsi používal internet, nebo ne?“

Ignoroval jsi někdy, co ti rodiče řekli, když jsi používal internet, nebo ne?

	Četnosti	Relativní četnosti
Ano, hodně	1253	6,7
Ano, trochu	5458	29,2
Ne	11058	59,1
Celkem platné	17769	95,0
Nevím	940	5,0
Celkem	18709	

Nyní předkládáme tabulku, obsahující regresní koeficienty našeho modelu, poměr šancí a některé další statistiky.

Tabulka 2

Regresní koeficienty a poměr šancí

	B	SE	Wald	df	p	poměr šancí
Konstanta	-0,08	0,23	0,13	1	0,72	0,92
Jak často používáš internet?						
Méně než 1x za měsíc (ref.)			1122,64	3	<0,001	
x Každý den/téměř každý den	-1,80	0,09	409,55	1	<0,001	0,17
x 1x/2x za týden	-0,58	0,09	39,22	1	<0,001	0,56
x 1x/2x za měsíc	0,02	0,14	0,02	1	<0,001	1,02
Myslíš, že jsou na internetu věci, které by lidi tvého věku jakkoliv obtěžovaly?						
Ano x ne	-0,15	0,04	13,17	1	<0,001	0,86
Jak moc si myslíš, že tví rodiče ví o tom, co děláš na internetu?						
Nic (ref.)			47,24	3	<0,001	
x Hodně	-0,16	0,09	1,77	1	0,183	0,89
x Docela dost	-0,29	0,08	12,09	1	0,001	0,75
x Trošku	-0,46	0,09	27,88	1	<0,001	0,63
Vím více o internetu, než mí rodiče						
Pravda (ref.)			177,66	2	<0,001	
x Není to pravda	0,76	0,06	171,76	1	<0,001	2,14
x Částečně pravda	0,27	0,05	26,05	1	<0,001	1,31
Vím hodně o používání internetu						
Pravda (ref.)			206,68	2	<0,001	
x Není to pravda	0,97	0,08	167,01	1	<0,001	2,64
x Částečně pravda	0,60	0,05	145,61	1	<0,001	1,83
Na internetu je mnoho věcí, prospěšných pro děti mého věku						
Pravda (ref.)			3,29	2	0,19	
x Není to pravda	-0,15	0,08	3,28	1	0,07	0,86
x Částečně pravda	-0,01	0,04	0,09	1	0,76	0,97
Ignoroval jsi někdy, co ti rodiče řekli, když jsi používal internet, nebo ne?						
Ne (ref.)			10,72	2	0,01	
x Ano, hodně	-0,20	0,09	5,46	1	0,02	0,82
x Ano, trochu	-0,12	0,05	7,09	1	0,01	0,88
Doporučoval ti někdy některý z tvých rodičů, jak používat internet bezpečně?						
Ano x ne	0,16	0,04	13,18	1	<0,001	1,17

Interpretace výsledků:

Po této analýze jsme zjistili, že náš model celkově signifikantně predikuje to, zda mají děti profil na sociálních sítích či ne ($R^2 = 0,324$ [Cox & Snell], 0,432

[Nagelkerke], $\chi^2(16, 15151) = 5925,96$, $p < 0,00$), přičemž i většina prediktorů se ukázala být signifikantními. Výjimku tvoří pouze dummy proměnné "Jak často používáte internet? - Jednou až dvakrát měsíčně", "Kolik myslíš, že toho rodiče vědí o tom, co děláš na internetu? - Hodně". Dále signifikantní nebyly všechny dummy proměnné u prediktoru "Myslíš, že na internetu je hodně dobrého pro děti tvého věku?", což znamená, že tato proměnná v žádné kategorii nepřispívá predikci.

Jako výrazný prediktor nám vyšla dummy proměnná "Vím více o internetu, než moji rodiče - Není to pravda". U dětí, které o internetu vědí méně než jejich rodiče, se šance, že nebude mít profil na sociální síti zvyšuje o 2,14 (pokud ve všech ostatních proměnných spadá do referenční kategorie). Ještě výraznější je dummy proměnná "Vím toho hodně o používání internetu - Není to pravda". U dětí, které toho dle svého mínění o používání internetu moc nevěděly, se šance, že nemají profil na sociální síti, zvyšuje o 2,64 (pokud ve všech ostatních proměnných spadá do referenční kategorie).

Náš model tedy vysvětluje zhruba 32% rozptylu proměnné (resp. 43%) „Máš v současnosti profil na sociální síti, který používáš, nebo ne?“, přičemž jako nejlepší prediktory se ukázaly být proměnné " Vím více o internetu, než moji rodiče - Není to pravda " a " Vím toho hodně o používání internetu - Není to pravda", z čehož vyplývá, že děti, které dle svého mínění neumí příliš používat internet ani ve srovnání se svými rodiči, budou mít méně pravděpodobně aktivně používaný profil na sociálních sítích.

Máte hezký model, mohli byste ho doopravdy interpretovat.