

Kontrasty v logistické regresi

BLANKA ŘEHÁKOVÁ*

Sociologický ústav AV ČR, v.v.i., Praha

Contrasts in Logistic Regression

Abstract: The article deals with the use of various types of contrasts, especially in logistic regression. Contrasts were originally developed within the framework of the analysis of variance. They gradually expanded into other statistical methods, for example, into logistic regression and loglinear or logit models. They are also used in linear regression when there are categorical variables among the explaining variables. Contrasts represent a method for working with variables that are not numerical but categorical. The article refers to the well-known types, such as indicator, simple, deviation, repeated, Helmert, difference and polynomial contrasts. Several others are also proposed. Contrasts are classified according to the appropriateness of their use for different types of categorical variables (nominal, ordinal). Their meaning and effect on the interpretation of odds ratios are explained on the basis of examples created using real data.

Keywords: logistic regression, odds ratios, contrast matrix, design matrix, quantitative methodology.

Sociologický časopis/Czech Sociological Review, 2008, Vol. 44, No. 4: 745–765

Cílem tohoto článku je připomenout základní pojmy logistické regrese, ale hlavně seznámit čtenáře s různými typy kontrastů a s jejich využíváním, v neposlední řadě pak s významem poměrů šancí při zvoleném typu kontrastů. Stať se neomezuje jen na nejužívanější typy kontrastů, jimiž jsou Indicator a Deviation – zmiňované někdy v literatuře pod názvy „dummy“ a „effect“ kódování –, protože by chtěla přispět k rozšíření spektra používaných typů kontrastů mezi českými sociology.

Dnes již klasickou knihou o logistické regresi je Hosmer, Lemeshow [1989], kde si čtenář může doplnit případné neznalosti. Kontrastům se hodně věnuje Bock [1975], ale jedná se o náročné čtení. Dalším zdrojem informací může být Menard [1995] nebo manuál SPSS [1999].¹ Problematika kontrastů se objevu-

* Veškerou korespondenci posílejte na adresu: RNDr. Blanka Řeháková, CSc., oddělení Hodnotové orientace ve společnosti, Sociologický ústav AV ČR, v.v.i., Jilská 1, 110 00 Praha 1, e-mail: Blanka.Rehakova@soc.cas.cz.

¹ Uvádím tento manuál jako zdroj pro tvar matice kontrastů všech typů, kterými se budeme zabývat, a to pro kategorizované proměnné s I kategoriemi, zatímco v tomto článku jsou uvedeny tyto matice jen pro případ I = 4. Uživatel je nepotřebuje znát, na druhou stranu se ale pomocí nich naučí, jak vytvářet matice svých vlastních kontrastů.

je v okamžiku, kdy je účelné porovnávat úroveň vysvětlované proměnné při změně kategorie nečíselné vysvětlující proměnné. Je totiž nepřípustné zahrnout nečíselné proměnné do modelu stejným způsobem jako číselné proměnné, protože čísla, která označují různé úrovně nečíselné proměnné, nemají numerický význam. Různé typy kontrastů byly původně vyvinuty v rámci analýzy rozptylu a odtud se rozšířily i do řady dalších statistických analytických postupů, například právě do logistické regrese nebo do logaritmicko-lineárních či logitových modelů.

Budeme uvažovat nejjednodušší případ logistické regrese, ve kterém vysvětlovaná proměnná Y nabývá pouze dvou hodnot, a sice 0 a 1. Budeme předpokládat, že $Y = 1$, jestliže se u respondenta objevil zkoumaný jev (je-li například Y účast ve volbách, pak $Y = 1$, jestliže se respondent zúčastnil voleb, $Y = 0$, jestliže se nezúčastnil). K vysvětlení proměnné Y máme řadu vysvětlujících proměnných $X_1, X_2, \dots, X_k$, které mohou být číselné nebo kategorizované (nečíselné), a to jak nominální, tak ordinální. Kdyby byla vysvětlovaná proměnná Y číselná, mohli bychom použít například lineární regresi

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k.$$

Analogem tohoto přístupu v případě binární (dvouhodnotové, dichotomické, alternativní) proměnné Y je logistická regrese

$$\text{logit}(Y) = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k,$$

kde $\text{logit}(Y)$ je definován jako $\ln \{P(Y = 1) / P(Y = 0)\}$. Logit není definován, pokud $P(Y = 1) = 0$ nebo $P(Y = 0) = 0$.²

Klíčovým pojmem v logistické regresi je šance (odds).

$$\begin{aligned} \text{šance}(Y = 1) &= \exp [\text{logit}(Y)] = \exp(B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k) \\ &= P(Y = 1) / P(Y = 0), \end{aligned}$$

kde $\exp$ označuje exponenciální funkci.³ Pravděpodobnost, šance a logit jsou základní pojmy logistické regrese a nesmí být v žádném případě směřovány.

² $\ln$ je přirozený logaritmus, tj. logaritmus při základu $e = 2,718$, P je pravděpodobnost jevu, na příklad jevu, že $Y = 1$. Připomeňme si, že pravděpodobnost je libovolné číslo mezi nulou a jedničkou, rovná se nule pro jev nemožný a rovná se jedničce pro jev jistý. Hodnoty logitu se pohybují od minus nekonečna do plus nekonečna.

³ Exponenciální funkce $\exp(x) = e^x$ je definovaná pro všechna reálná x , je vždy kladná, je rostoucí na celém svém definičním oboru, je menší než jedna pro všechna $x < 0$, je větší než jedna pro všechna $x > 0$, $\exp(0) = 1$. Dále platí tyto vztahy: $\exp(x + y) = \exp(x) \exp(y)$, $\exp(\Sigma x_i) = \Pi(\exp(x_i))$, $\exp(x - y) = \exp(x) / \exp(y)$, $[\exp(x)]^y = \exp(xy)$, $\exp(-x) = 1 / \exp(x)$. Pro $x > 0$ platí $\exp(\ln x) = x$.

V odborné literatuře se nejčastěji interpretují šance a jejich poměry. $P(Y = 1)$, šance($Y = 1$) a $\text{logit}(Y)$ jsou spojeny vztahem

$$P(Y = 1) = \text{šance}(Y = 1) / [1 + \text{šance}(Y = 1)] = 1 / \{\exp[-\text{logit}(Y)] + 1\}.$$

Problematickou použití kontrastů a s tím spojenou interpretaci parametrů budeme demonstrovat na příkladech, které byly vytvořeny pomocí procedury Logistic Regression z SPSS na části souboru dat z výzkumu Naše společnost 2006, který provedlo CVVM ve dnech 9.–21. 6. 2006. Do zpracování byli zahrnuti jen ti respondenti, kteří volili ve volbách do Parlamentu České republiky v roce 2006 buď ODS, nebo ČSSD.

Vysvětlovanou proměnnou je Volba. Volba = 1, jestliže respondent prohlásil, že volil ODS, Volba = 0, jestliže respondent nevolil ODS. Protože soubor obsahuje jen voliče ODS a ČSSD, znamená 0 současně volbu ČSSD. Vysvětlující proměnnou je nejvyšší dosažené vzdělání respondenta (Edu). Je to ordinální kategorizovaná proměnná, která má čtyři kategorie: 1 = základní vzdělání, 2 = další vzdělání bez maturity, 3 = další vzdělání s maturitou, 4 = pomaturitní vzdělání. Číslice 1, 2, 3, 4 zde nemají význam čísel, slouží jen k označení jednotlivých kategorií vzdělání.

Pokud by proměnná Edu byla číselná (na příklad počet let studia), byla by rovnice logistické regrese $\text{logit}(\text{Volba}) = B_0 + B_1 \text{Edu}$. Proměnná Edu ale číselná není a s tím se musíme nějak vypořádat. Jako řešení se nabízí analogické opatření, které se používá například v lineární regresi. Každá vysvětlující proměnná, která není číselná, se nahradí umělými proměnnými, které Hosmer a Lemeshow [Hosmer, Lemeshow 1989] nazývají designové proměnné. Snahou je reprezentovat každou kategorii jednou designovou proměnnou. To však není možné. Pokud je v rovnici logistické regrese s jednou vysvětlující proměnnou, která je kategorizovaná a která má I kategorií, konstanta B_0 , je možné ze vztahů, které máme k dispozici, odhadnout už jen dalších $I - 1$ parametrů $B_1, B_2, \dots, B_{I-1}$. Z toho důvodu lze reprezentovat kategorizovanou proměnnou s I kategoriemi pouze $I - 1$ designovými proměnnými. Jaké proměnné to budou, tj. jakých hodnot budou nabývat, úzce souvisí s typem kontrastů, které určují význam parametrů B. Proměnnou Edu, která má čtyři kategorie, budou reprezentovat tři nové proměnné Edu(1), Edu(2), Edu(3) a regresní rovnice bude mít tvar

$$\text{logit}(\text{Volba}) = B_0 + B_1 \text{Edu}(1) + B_2 \text{Edu}(2) + B_3 \text{Edu}(3).$$

Rozšíření na situaci, kdy vysvětlující proměnná má více než čtyři kategorie, nebo když má jen tři kategorie, se konceptuálně neliší.⁴

Procedura Logistic Regression nabízí tyto typy kontrastů: Indicator (refcat), Simple (refcat), Deviation (refcat), Repeated, Helmert, Difference, Polynomi-

⁴ Hosmer a Lemeshow [Hosmer, Lemeshow 1989] doporučují, aby vysvětlující proměnné, které mají jen dvě kategorie, byly kódovány nebo překódovány pro analýzu pomocí nul a jedniček a aby se s nimi v logistické regresi pracovalo jako s číselnými.

al (metric).⁵ Kromě těchto kontrastů si může uživatel vytvořit své vlastní. Výběr kontrastů by neměl být samoučelný, ale měl by odpovídat tomu, co chceme zjistit, respektive jakou hypotézu chceme ověřit. Měl by být též v souladu s typem proměnné (nominální nebo ordinální). Výběr kontrastů nemá vliv na hodnoty logitů, šancí, pravděpodobností, ani na ukazatele kvality modelu. Ty jsou na kontrastech nezávislé. Pro různé typy kontrastů ale budeme dostávat různé hodnoty parametrů B , a tudíž i $\exp(B)$, neboť význam parametrů souvisí s typem kontrastů, jak bude zřejmé z další části stati.

Kontrasty typu Indicator (refcat)

Tento typ kontrastů bývá používán nejčastěji, a to i v případech, kdy nejsou pro interpretaci optimální. Vhodný je tehdy, když se zajímáme o porovnání efektů různých kategorií s efektem jedné další, kterou nazýváme referenční (refcat). Lze ho použít jak pro nominální, tak pro ordinální proměnné, i když pro ordinální proměnné jsou vhodnější jiné typy (Repeated, Polynomial). Může se ovšem stát, že ordinální proměnná se v dané úloze neprojevuje ordinálně,⁶ pak je typ Indicator zcela na místě. Když zvolíme za referenční kategorii základní vzdělání, tj. první kategorii, bude mít matice kontrastů tvar, který je v tabulce 1.⁷

První řádek v tabulce 1 indikuje význam parametru B_0 (konstanty) jako efekt první kategorie, tj. základního vzdělání. Druhý řádek indikuje význam parametru B_1 jako porovnání efektu druhé a první kategorie, tj. dalšího vzdělání bez maturity a základního vzdělání. Třetí řádek indikuje význam parametru B_2 jako porovnání efektu třetí a první kategorie, tj. dalšího vzdělání s maturitou a základního vzdělání. Čtvrtý řádek indikuje význam parametru B_3 jako porovnání efektu čtvrté a první kategorie, tj. pomaturitního a základního vzdělání.

Druhý, třetí a čtvrtý sloupec designové matice, která je v tabulce 2, ukazují, jak jsou v případě kontrastů Indicator(1) kódované designové proměnné Edu(1), Edu(2), Edu(3).⁸ První sloupec se vztahuje ke konstantě. Proměnná Edu(1) se rov-

⁵ Názvy Simple, Deviation, Helmert, Polynomial používá pro stejné typy kontrastů Bock [1975]. Jiné typy ve své knize nepoužívá. Pro typy Simple, Deviation a Helmert uvádí jak designové matice, které nazývá „basis“, tak matice kontrastů, a to pro proměnné s I kategoriemi. V různých programových systémech se můžeme setkat s různými názvy kontrastů, proto je vždy dobré si ověřit, jaký typ se pod použitým názvem skrývá, aby byla interpretace výsledků správná.

⁶ Vysvětlení předvedu na příkladu: Jestliže efekty kategorií vzdělání rostou (resp. klesají) s růstem vzdělání, pak se proměnná Edu chová ordinálně. Jestliže se rostoucí či klesající trend v efektech neprojevuje, Edu se v dané úloze nechová ordinálně, i když sama ordinální je, neboť její kategorie jsou uspořádané.

⁷ Kontrasty jsou v druhém, třetím a čtvrtém řádku. První řádek udává význam konstanty B_0 nejen v této matici kontrastů, ale i ve všech dalších.

⁸ Kódování designových proměnných je v Logistic Regression součástí výstupu pod hlavičkou Categorical Variables Codings.

Tabulka 1. Matice kontrastů Indicator(1)

1	0	0	0
-1	1	0	0
-1	0	1	0
-1	0	0	1

Zdroj: autorka.

Tabulka 2. Designová matice (Indicator(1))

1	0	0	0
1	1	0	0
1	0	1	0
1	0	0	1

Zdroj: výstup z Logistic Regression.

Tabulka 3. Odhady parametrů a testy významnosti (Indicator(1))

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	,588	,290	4,096	1	,043	1,800
Edu(2)	1,472	,293	25,277	1	,000	4,359
Edu(3)	1,724	,310	30,979	1	,000	5,607
Konstanta	-,990	,265	13,968	1	,000	,371

Zdroj: Naše společnost 2006.

ná jedné pro respondenty s dalším vzděláním bez maturity a pro respondenty s jinými typy vzdělání se rovná nule. Proměnná Edu(2) se rovná jedné pro respondenty s dalším vzděláním s maturitou a pro respondenty s jiným typem vzdělání se rovná nule. Proměnná Edu(3) se rovná jedné pro respondenty s pomaturitním vzděláním a nule pro respondenty s jinými typy vzdělání. Designová matice a matice kontrastů jsou k sobě navzájem inverzní,⁹ tj. jejich součin (v libovolném pořadí) se rovná jednotkové matici.¹⁰

V tabulce 3 jsou odhady parametrů (B), jejich směrodatné chyby (S.E.), Waldova statistika pro test, že se příslušný parametr rovná nule (Wald), stupně volnosti (df), dosažená hladina významnosti (Sig.) a exp(B). Rovnice modelu je

$$\text{logit}(\text{Volba}) = -0,990 + 0,588 \text{ Edu}(1) + 1,472 \text{ Edu}(2) + 1,724 \text{ Edu}(3).$$

Odtud lze spočítat logity a následovně šance a pravděpodobnosti pro jednotlivé kategorie vzdělání tak, že za Edu(1), Edu(2), Edu(3) dosazujeme příslušné

⁹ Designová matice a matice kontrastů jsou inverzní matice i ve všech dalších případech zde uvedených.

¹⁰ Jednotková matice je čtvercová matice, tj. matice, která má stejný počet řádků a sloupců a která má na diagonále jedničky a všude jinde nuly.

hodnoty z tabulky 2. Přesně řečeno jedná se o odhady logitů, šancí a pravděpodobností, ale jelikož zde pracujeme jen s odhady, dovolíme si slůvko „odhad“ neuvádět.

Pro základní vzdělání

$$\begin{aligned}\text{logit}(\text{Volba}) &= -0,990, \text{šance}(\text{Volba} = 1) = 0,371, P(\text{Volba} = 1) = \\ &= 0,371 / (1 + 0,371) = 1 / (1 + \exp(0,990)) = 0,271.\end{aligned}$$

Pro další vzdělání bez maturity

$$\begin{aligned}\text{logit}(\text{Volba}) &= -0,402, \text{šance}(\text{Volba} = 1) = 0,669, P(\text{Volba} = 1) = \\ &= 0,669 / (1 + 0,669) = 1 / (1 + \exp(0,402)) = 0,401.\end{aligned}$$

Pro další vzdělání s maturitou

$$\begin{aligned}\text{logit}(\text{Volba}) &= 0,482, \text{šance}(\text{Volba} = 1) = 1,619, P(\text{Volba} = 1) = \\ &= 1,619 / (1 + 1,619) = 1 / (1 + \exp(-0,482)) = 0,618.\end{aligned}$$

Pro pomaturitní vzdělání

$$\begin{aligned}\text{logit}(\text{Volba}) &= 0,734, \text{šance}(\text{Volba} = 1) = 2,083, P(\text{Volba} = 1) = \\ &= 2,083 / (1 + 2,083) = 1 / (1 + \exp(-0,734)) = 0,676.\end{aligned}$$

V publikacích se obvykle interpretují parametry B nebo $\exp(B)$, kterým se říká poměry šancí (odds ratios). Ty jsou uvedeny v posledním sloupci tabulky 3. O jaké poměry se ale jedná? O takové, jaké kontrasty jsme použili. Z definice kontrastů Indicator(1) plyne, že B_1 (hodnota ve sloupci B a v řádku Edu(1) tabulky 3) je rozdíl logitů pro další vzdělání bez maturity a základní vzdělání, tj. $-0,402 - (-0,990) = 0,588$ a $\exp(B_1) = \exp(0,588) = 1,800$ je poměr šancí volby ODS pro voliče s dalším vzděláním bez maturity a se základním vzděláním. B_2 (hodnota ve sloupci B a v řádku Edu(2) tabulky 3) je rozdíl logitů pro další vzdělání s maturitou a základní vzdělání, tj. $0,482 - (-0,990) = 1,472$ a $\exp(B_2) = \exp(1,472) = 4,359$ je poměr šancí volby ODS pro voliče s dalším vzděláním s maturitou a se základním vzděláním. B_3 (hodnota ve sloupci B a v řádku Edu(3) tabulky 3) je rozdíl logitů pro pomaturitní vzdělání a základní vzdělání, tj. $0,734 - (-0,990) = 1,724$ a $\exp(B_3) = \exp(1,724) = 5,607$ je poměr šancí volby ODS pro voliče s pomaturitním vzděláním a se základním vzděláním. Konstanta B_0 se rovná $\text{logit}(\text{Volba})$ pro základní vzdělání, tj. pro referenční kategorii.

Kontrasty typu Simple (refcat)

Naprosto stejnou službu nám udělají kontrasty Simple (refcat), neboť rovněž ony se používají v situaci, kdy se zajímáme o efekty kategorií kategorizované proměnné vzhledem k referenční kategorii. A to vše navzdory tomu, že proměnné Edu(1), Edu(2), Edu(3) jsou kódované zcela jinak, i když referenční kategorie je opět základní vzdělání, jak se můžeme přesvědčit porovnáním tabulek 1 a 4. Matice kontrastů Simple(1) je v tabulce 4. Samotné kontrasty jsou v druhém, třetím a čtvrtém řádku, první indikuje význam konstanty. Právě a pouze ten je jiný než u matice kontrastů typu Indicator. U kontrastů typu Simple (a všech dalších, které budou ještě následovat) se konstanta B_0 rovná průměrnému efektu kategorií proměnné Edu, tedy průměru ze všech čtyř logitů.

Z druhého, třetího a čtvrtého sloupce tabulky 5 je zřejmé, že proměnná Edu(1) se nyní rovná $(-1/4)$ pro respondenty se základním vzděláním, s dalším vzděláním s maturitou a s pomaturitním vzděláním a $3/4$ pro respondenty s dalším vzděláním bez maturity. Proměnná Edu(2) se rovná $3/4$ pro respondenty s dalším vzděláním s maturitou a $(-1/4)$ v ostatních případech. Proměnná Edu(3) se rovná $3/4$ pro respondenty s pomaturitním vzděláním a $(-1/4)$ v ostatních případech. První sloupec se vztahuje ke konstantě.

Jak už bylo řečeno, význam parametrů B_1 , B_2 , B_3 , a tudíž i $\exp(B_1)$, $\exp(B_2)$,

Tabulka 4. Matice kontrastů Simple(1)

1/4	1/4	1/4	1/4
-1	1	0	0
-1	0	1	0
-1	0	0	1

Zdroj: autorka podle manuálu SPSS [1999].*

Tabulka 5. Designová matice (Simple(1))

1	-1/4	-1/4	-1/4
1	3/4	-1/4	-1/4
1	-1/4	3/4	-1/4
1	-1/4	-1/4	3/4

Zdroj: výstup z Logistic Regression.

Tabulka 6. Odhady parametrů a testy významnosti (Simple(1))

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	,588	,290	4,096	1	,043	1,800
Edu(2)	1,472	,293	25,277	1	,000	4,359
Edu(3)	1,724	,310	30,979	1	,000	5,607
Konstanta	-,044	,089	,251	1	,617	,957

Zdroj: Naše společnost 2006.

* Další tabulky uváděné bez zdroje mají stejný zdroj jako tabulka 4.

$\exp(B_3)$ je stejný jako v případě kontrastů Indicator(1), konstanta B_0 je ale průměrný efekt všech kategorií, nikoliv efekt referenční kategorie, jak tomu je u kontrastů typu Indicator (refcat). Tabulka 6 se liší od tabulky 3 právě jen v posledním řádku.

Kontrasty typu Deviation (refcat)

Pokud se zajímáme o efekty jednotlivých kategorií vzhledem k průměrnému efektu všech kategorií, zvolíme kontrasty Deviation (refcat). Tyto kontrasty jsou vhodné jak pro nominální, tak pro ordinální proměnné. Za referenční kategorii zvolíme tentokrát pro změnu čtvrtou, tj. pomaturitní vzdělání. Matice kontrastů Deviation(4) je v tabulce 7, designová matice je v tabulce 8, výsledky jsou v tabulce 9. V tabulce 9 chybí parametr B s indexem shodným s číslem referenční kategorie, v našem případě B_4 . Dopočítáme ho pomocí rovnice $B_4 = 0 - (B_1 + B_2 + B_3) = 0,778$. Kdyby byl index referenční kategorie 1, chyběl by parametr B_1 , a ve výsledné tabulce by se objevily parametry B v pořadí B_2, B_3, B_4 . B_1 bychom dopočítali ze vztahu $B_1 = 0 - (B_2 + B_3 + B_4)$. Kdyby byl index referenční kategorie 3, chyběl by parametr B_3 a ve výsledné tabulce by se objevily parametry B v pořadí B_1, B_2, B_4 . B_3 bychom dopočítali ze vztahu $B_3 = 0 - (B_1 + B_2 + B_4)$. Obdobně pro referenční kategorii 2.

Význam parametrů B a $\exp(B)$ je dán definicí kontrastů Deviation. Podle ní jsou B_1, B_2, B_3, B_4 efekty jednotlivých kategorií (tj. logity pro jednotlivé kategorie) vzhledem k průměrnému efektu všech kategorií (tj. průměru všech logitů). Můžeme se o tom snadno přesvědčit přímým výpočtem.

$$B_1 = -0,990 - 1/4 (-0,990 - 0,402 + 0,482 + 0,734) = -0,946,$$

$$B_2 = -0,402 - 1/4 (-0,990 - 0,402 + 0,482 + 0,734) = -0,358,$$

$$B_3 = 0,482 - 1/4 (-0,990 - 0,402 + 0,482 + 0,734) = 0,526,$$

$$B_4 = 0,734 - 1/4 (-0,990 - 0,402 + 0,482 + 0,734) = 0,778.$$

Číslo $\exp(B_i)$ má význam podílu šance volby ODS voliči se základním vzděláním a geometrickým průměrem šancí volby ODS voliči se základním vzděláním, s dalším vzděláním bez maturity, s dalším vzděláním s maturitou a s pomaturitním vzděláním. Platí totiž, že

$$\begin{aligned} \exp(B_1) &= \exp\{-0,990 - 1/4 (-0,990 - 0,402 + 0,482 + 0,734)\} = \\ &= \exp(-0,990) / \exp[1/4 (-0,990 - 0,402 + 0,482 + 0,734)] = \\ &= \exp(-0,990) / \{\exp(-0,990) \exp(-0,402) \exp(0,482) \exp(0,734)\}^{1/4}. \end{aligned}$$

Analogicky pro $\exp(B_2)$, $\exp(B_3)$ a $\exp(B_4)$.¹¹

¹¹ Geometrický průměr čísel $k_1, k_2, \dots, k_n$ se rovná n -té odmocnině ze součinu těchto čísel, tj. $(k_1 k_2 \dots k_n)^{1/n}$.

Tabulka 7. Matice kontrastů Deviation(4) Tabulka 8. Designová matice (Deviation(4))

$\begin{vmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 3/4 & -1/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ -1/4 & -1/4 & 3/4 & -1/4 \end{vmatrix}$	$\begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{vmatrix}$
---	---

Zdroj: výstup z Logistic Regression.

Tabulka 9. Odhady parametrů a testy významnosti (Deviation(4))

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	-,946	,207	20,832	1	,000	,388
Edu(2)	-,358	,122	8,603	1	,003	,699
Edu(3)	,526	,125	17,735	1	,000	1,693
Konstanta	-,044	,089	,251	1	,617	,957

Zdroj: Naše společnost 2006.

Kontrasty typu Repeated

Kontrasty Repeated jsou vhodné pro ordinální proměnné (tabulka 10, 11). Porovnávají efekt první kategorie s efektem druhé, efekt druhé s efektem třetí atd. Pomocí kontrastů Repeated můžeme ověřit hypotézu, že se snižující se úrovní vzdělání šance volby ODS klesají. Z výsledků tabulky 12 vidíme, že to není pravda. Parametry B jsou sice všechny záporné, a tudíž hodnoty $\exp(B)$ jsou menší než jedna, ale významně odlišné od nuly jsou jen první dva. Rozdíl efektů maturitního a pomaturitního vzdělání není významný.

Podle definice kontrastů Repeated mají parametry B_1 , B_2 , B_3 tento význam:

B_1 je rozdíl logitů pro základní vzdělání a další vzdělání bez maturity, tj. efekt první kategorie vzhledem ke druhé

$$B_1 = -0,990 - (-0,402) = -0,588.$$

B_2 je rozdíl logitů pro další vzdělání bez maturity a další vzdělání s maturitou, tj. efekt druhé kategorie vzhledem ke třetí

$$B_2 = -0,402 - 0,482 = -0,884.$$

B_3 je rozdíl logitů pro další vzdělání s maturitou a pomaturitní vzdělání, tj. efekt třetí kategorie vzhledem ke čtvrté

$$B_3 = 0,482 - 0,734 = -0,252.$$

Konstanta je průměr všech logitů, tj. průměrný efekt všech kategorií vzdělání. $\exp(B_1)$ je poměr šancí volby ODS voliči se základním vzděláním a s dalším vzděláním bez maturity, $\exp(B_2)$ je poměr šancí volby ODS voliči s dalším vzděláním bez maturity a s dalším vzděláním s maturitou, $\exp(B_3)$ je poměr šancí volby ODS voliči s dalším vzděláním s maturitou a s pomaturitním vzděláním.

Pokud bychom raději znali efekty druhé kategorie vzhledem k první, třetí vzhledem k druhé a čtvrté vzhledem ke třetí, můžeme použít kontrasty Repeated, ale musíme změnit znaménka u parametrů B a pro tyto nové hodnoty dopočítat $\exp(B)$, nebo použít převrácené hodnoty $\exp(B)$ z tabulky 12, tj. $1 / \exp(B)$. Ostatní sloupce tabulky 12 zůstávají v platnosti beze změn. Pak je tu ještě možnost, že vytvoříme kontrasty, které budou odpovídat našemu požadavku. Jejich matice se rovná (-1) násobku matice kontrastů Repeated, která je v tabulce 10 (první řádek se ovšem nemění). Matici nových kontrastů tedy dostaneme tak, že každý prvek matice kontrastů Repeated, kromě prvků v prvním řádku, vynásobíme číslem (-1) . Speciálním kontrastům se budeme věnovat podrobněji v závěrečné části této stati.

Tabulka 10. Matice kontrastů Repeated

1/4	1/4	1/4	1/4
1	-1	0	0
0	1	-1	0
0	0	1	-1

Tabulka 11. Designová matice (Repeated)

1	3/4	1/2	1/4
1	-1/4	1/2	1/4
1	-1/4	-1/2	1/4
1	-1/4	-1/2	-3/4

Zdroj: výstup z Logistic Regression.

Tabulka 12. Odhady parametrů a testy významnosti (Repeated)

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	-,588	,290	4,096	1	,043	,556
Edu(2)	-,884	,172	26,374	1	,000	,413
Edu(3)	-,252	,203	1,536	1	,215	,777
Konstanta	-,044	,089	,251	1	,617	,957

Zdroj: Naše společnost 2006.

Kontrasty typu Helmert

Helmertovy kontrasty jsou zajímavé v případech, kdy existuje apriorní logické uspořádání kategorií. Porovnávají efekt první kategorie s průměrným efektem všech následujících kategorií, efekt druhé kategorie s průměrným efektem všech následujících kategorií, atd. až nakonec efekt předposlední kategorie s efektem poslední kategorie (viz tabulku 13). Použití těchto kontrastů pro proměnnou Edu se mi nejeví jako smysluplné. Uvažujme místo ní proměnnou, kterou označíme Empl. Je to kategorizovaná proměnná o čtyřech kategoriích, přičemž do první kategorie patří ti, kteří nejsou výdělečně činní, do druhé patří samostatně výdělečně činní, do třetí zaměstnanci soukromého sektoru, do čtvrté ostatní zaměstnanci. Vysvětlovanou proměnnou bude opět Volba. Helmertovy kontrasty umožní porovnání efektu těch, kteří nejsou výdělečně činní, s průměrným efektem tří typů výdělečně činných, porovnání efektu samostatně výdělečně činných s průměrným efektem zaměstnanců v soukromém sektoru a ostatních zaměstnanců a konečně porovnání efektu zaměstnanců v soukromém sektoru s efektem ostatních zaměstnanců. Designová matice je v tabulce 14.

Rovnice modelu je $\text{logit}(\text{Volba}) = 0,302 - 0,707 \text{ Empl}(1) + 1,335 \text{ Empl}(2) - 0,084 \text{ Empl}(3)$. Z ní určíme logity, šance a pravděpodobnosti.

Pro nepracující

$$\text{logit}(\text{Volba}) = -0,228, \text{ šance}(\text{Volba} = 1) = 0,796, P(\text{Volba} = 1) = 0,443.$$

Pro podnikatele

$$\text{logit}(\text{Volba}) = 1,369, \text{ šance}(\text{Volba} = 1) = 3,931, P(\text{Volba} = 1) = 0,797.$$

Pro zaměstnance v soukromém sektoru

$$\text{logit}(\text{Volba}) = -0,008, \text{ šance}(\text{Volba} = 1) = 0,992, P(\text{Volba} = 1) = 0,498.$$

Pro zaměstnance v jiném sektoru

$$\text{logit}(\text{Volba}) = 0,076, \text{ šance}(\text{Volba} = 1) = 1,079, P(\text{Volba} = 1) = 0,519.$$

Podle definice kontrastů Helmert je parametr B_1 rozdíl efektu první kategorie proměnné Empl a průměrného efektu dalších třech kategorií, B_2 je rozdíl efektu druhé kategorie a průměrného efektu třetí a čtvrté kategorie, B_3 je rozdíl efektu třetí a čtvrté kategorie. B_0 je průměrný efekt všech kategorií.

$$B_1 = -0,228 - 1/3(1,369 - 0,008 + 0,076) = -0,707,$$

Tabulka 13. Matice kontrastů Helmert

1/4	1/4	1/4	1/4
1	-1/3	-1/3	-1/3
0	1	-1/2	-1/2
0	0	1	-1

Tabulka 14. Designová matice (Helmert)

1	3/4	0	0
1	-1/4	2/3	0
1	-1/4	-1/3	1/2
1	-1/4	-1/3	-1/2

Zdroj: výstup z Logictic Regression.

Tabulka 15. Odhady parametrů a testy významnosti (Helmert)

	B	S.E.	Wald	df	Sig.	Exp(B)
Empl			37,064	3	,000	
Empl(1)	-,707	,160	19,532	1	,000	,493
Empl(2)	1,335	,256	27,172	1	,000	3,799
Empl(3)	-,084	,206	,166	1	,684	,920
Konstanta	,302	,084	13,022	1	,000	1,353

Zdroj: Naše společnost 2006.

$$B_2 = 1,369 - 1/2(-0,008 + 0,076) = 1,335,$$

$$B_3 = -0,008 - 0,076 = -0,084,$$

$$B_0 = 1/4(-0,228 + 1,369 - 0,008 + 0,076) = 0,302.$$

Exp(B_1) je poměr šance volby ODS nepracujícími a geometrickým průměrem šancí volby ODS pracujícími, exp(B_2) je poměr šance volby ODS podnikateli a geometrickým průměrem šancí volby ODS zaměstnanci, exp(B_3) je poměr šancí volby ODS zaměstnanci v soukromém sektoru a zaměstnanci v jiném sektoru. Z tabulky 15 plyne, že nevýznamný je pouze rozdíl efektů zaměstnanců v soukromém a jiném sektoru. Šance volby ODS nepracujícími je 0,493krát menší, než činí geometrický průměr šancí volby ODS třemi skupinami pracujících. Šance volby ODS podnikateli je 3,799krát větší, než je geometrický průměr šancí volby ODS dvěma skupinami zaměstnanců.

Kontrasty typu Difference

Kontrasty typu Difference jsou obrácené Helmertovy kontrasty. Porovnávají efekt druhé a první kategorie, efekt třetí kategorie s průměrným efektem první a druhé kategorie, efekt čtvrté kategorie s průměrným efektem první, druhé a třetí kategorie, až nakonec efekt poslední kategorie s průměrným efektem všech předchozích kategorií (viz tabulku 16). V případě proměnné Edu porovnávají efekt dalšího vzdělání bez maturity s efektem základního vzdělání, efekt dalšího vzdělání s maturitou s průměrným efektem dvou typů vzdělání bez maturity a efekt pomaturitního vzdělání s průměrným efektem tří nižších typů vzdělání. Designová matice je v tabulce 17.

Podle definice kontrastů Difference je parametr B_1 rozdíl logitů pro další vzdělání bez maturity a základní vzdělání, tj. efekt druhé kategorie vzhledem k první

$$B_1 = -0,402 - (-0, 990) = 0,588.$$

Parametr B_2 je rozdíl logitu pro další vzdělání s maturitou a průměru logitů pro základní vzdělání a další bez maturity, tj. efekt třetí kategorie vzhledem k průměrnému efektu první a druhé kategorie

$$B_2 = 0,482 - 1/2 (-0, 990 - 0,402) = 1,178.$$

Parametr B_3 je rozdíl logitu pro pomaturitní vzdělání a průměru logitů pro základní vzdělání, další bez maturity a další s maturitou, tj. efekt čtvrté kategorie vzhledem k průměrnému efektu první, druhé a třetí kategorie

$$B_3 = 0,734 - 1/3 (-0, 990 - 0,402 + 0,482) = 1,037.$$

Číslo $\exp(B_1)$ je poměr šancí volby ODS voliči s dalším vzděláním bez maturity a se základním vzděláním. Číslo $\exp(B_2)$ je poměr šance volby ODS voliči s dalším vzděláním s maturitou a geometrickým průměrem šancí volby ODS

Tabulka 16. Matice kontrastů Difference

1/4	1/4	1/4	1/4
-1	1	0	0
-1/2	-1/2	1	0
-1/3	-1/3	-1/3	1

Tabulka 17. Designová matice (Difference)

1	-1/2	-1/3	-1/4
1	1/2	-1/3	-1/4
1	0	2/3	-1/4
1	0	0	3/4

Zdroj: výstup z Logistic Regression.

Tabulka 18. Odhady parametrů a testy významnosti (Difference)

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	,588	,290	4,096	1	,043	1,800
Edu(2)	1,178	,191	37,921	1	,000	3,249
Edu(3)	1,037	,192	29,224	1	,000	2,822
Konstanta	-,044	,089	,251	1	,617	,957

Zdroj: Naše společnost 2006.

voliči se základním vzděláním a s dalším vzděláním bez maturity. Číslo $\exp(B_3)$ je poměr šance volby ODS voliči s pomaturitním vzděláním a geometrickým průměrem šancí volby ODS voliči s nižšími typy vzdělání (viz tabulku 18).

Kontrasty typu Polynomial

Kontrasty typu Polynomial jsou vhodné pro ordinální kategorizované proměnné. Vycházejí z ortogonálních polynomů a jsou vytvořeny tak, aby parametr B_1 představoval lineární efekt všech kategorií, B_2 kvadratický, B_3 kubický atd. Matice polynomiálních kontrastů pro proměnnou se čtyřmi stejně vzdálenými kategoriemi je v tabulce 19, designová matice je v tabulce 20, výsledky jsou v tabulce 21.

Z tabulky 21 sloupce Sig. plyne, že významný je pouze lineární efekt. Pro parametry B_1 , B_2 , B_3 podle definice kontrastů Polynomial platí, že

$$B_1 = -3/\sqrt{20} (-0,990) - 1/\sqrt{20} (-0,402) + 1/\sqrt{20} (0,482) + 3/\sqrt{20} (0,734) = 1,354,$$

$$B_2 = 1/2 (-0,990) - 1/2 (-0,402) - 1/2 (0,482) + 1/2 (0,734) = -0,168,$$

$$B_3 = -1/\sqrt{20} (-0,990) + 3/\sqrt{20} (-0,402) - 3/\sqrt{20} (0,482) + 1/\sqrt{20} (0,734) = -0,208.$$

Abychom si přiblížili smysl a význam kontrastů Polynomial, budeme uvažovat ideální případ čistého lineárního vztahu mezi logity. Předpokládejme, že místo hodnot $(-0,990)$, $(-0,402)$, $0,482$, $0,734$ máme hodnoty $(-0,99)$, $(-0,39)$, $0,21$, $0,81$. Rozdíly mezi druhou a první hodnotou, třetí a druhou, čtvrtou a třetí jsou stejné, rovnají se $0,60$. Body $(1; -0,99)$, $(2; -0,39)$, $(3; 0,21)$, $(4; 0,81)$ leží na přímce $y = 0,60x - 1,59$. Když pro tyto nové logity spočteme hodnoty B_1 , B_2 , B_3 , zjistíme, že

$$B_1 = -3/\sqrt{20} (-0,99) - 1/\sqrt{20} (-0,39) + 1/\sqrt{20} (0,21) + 3/\sqrt{20} (0,81) = 1,342,$$

$$\exp(B_1) = 3,827,$$

Tabulka 19. Matice kontrastů Polynomial (stejně vzdálené kategorie)

1/4	1/4	1/4	1/4
$-3/\sqrt{20}$	$-1/\sqrt{20}$	$1/\sqrt{20}$	$3/\sqrt{20}$
1/2	-1/2	-1/2	1/2
$-1/\sqrt{20}$	$3/\sqrt{20}$	$-3/\sqrt{20}$	$1/\sqrt{20}$

Tabulka 20. Designová matice (Polynomial, stejně vzdálené kategorie)

1	$-3/\sqrt{20}$	1/2	$-1/\sqrt{20}$
1	$-1/\sqrt{20}$	-1/2	$3/\sqrt{20}$
1	$1/\sqrt{20}$	-1/2	$-3/\sqrt{20}$
1	$3/\sqrt{20}$	1/2	$1/\sqrt{20}$

Zdroj: výstup z Logistic Regression.

Tabulka 21. Odhady parametrů a testy významnosti (Polynomial, stejně vzdálené kategorie)

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	1,354	,211	41,069	1	,000	3,874
Edu(2)	-,168	,177	,899	1	,343	,845
Edu(3)	-,208	,135	2,380	1	,123	,812
Konstanta	-,044	,089	,251	1	,617	,957

Zdroj: Naše společnost 2006.

$$B_2 = 1/2 (-0,99) - 1/2 (-0,39) - 1/2 (0,21) + 1/2 (0,81) = 0, \exp(B_2) = 1,$$

$$B_3 = -1/\sqrt{20} (-0,99) + 3/\sqrt{20} (-0,39) - 3/\sqrt{20} (0,21) + 1/\sqrt{20} (0,81) = 0, \\ \exp(B_3) = 1.$$

Přítomný je jen lineární efekt všech kategorií, kvadratický a kubický jsou nulové. Pro šance $\exp(-0,99) = 0,372$, $\exp(-0,39) = 0,677$, $\exp(0,21) = 1,234$, $\exp(0,81) = 2,248$ to znamená, že poměry druhé a první, třetí a druhé, čtvrté a třetí jsou stejné, zde rovné $1,82 = \exp(0,60)$. Lze dokázat i opačné tvrzení: Když pro

Tabulka 22. Designová matice (Polynomial, nestejně vzdálené kategorie)

1	-,632	,500	-,316
1	-,316	-,500	,632
1	,316	-,500	-,632
1	,632	,500	,316

Zdroj: výstup z Logistic Regression.

Tabulka 23. Odhady parametrů a testy významnosti (Polynomial, nestejně vzdálené kategorie)

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	1,370	,203	45,401	1	,000	3,963
Edu(2)	-,168	,177	,899	1	,343	,845
Edu(3)	-,014	,146	,009	1	,923	,986
Constant	-,044	,089	,251	1	,617	,957

Zdroj: Naše společnost 2006.

nějaké logity L_1 , L_2 , L_3 , L_4 jsou B_2 a B_3 rovny nule, pak body (1, L_1), (2, L_2), (3, L_3), (4, L_4) leží na přímce.

Kontrasty Polynomial umožňují rovněž zahrnutí předpokladu, či apriorní znalosti o nestejně vzdálených kategoriích kategorizované proměnné. Předpokládejme, že vzdálenost mezi druhou a třetí kategorií proměnné Edu je dvojnásobkem vzdálenosti mezi první a druhou, třetí a čtvrtou. To vyjádříme metrikou (1, 2, 4, 5). Záleží pouze na relativních rozdílech mezi čísly uvedenými v závorce. Stejná metrika jako (1, 2, 4, 5) je například (1, 3, 7, 9) nebo (20, 30, 50, 60). Kódování proměnných Edu(1), Edu(2), Edu(3) se změní (porovnejte tabulky 20 a 22), změní se pochopitelně i výsledky (porovnejte tabulky 21 a 23). Lineární efekt kategorií se prakticky nezměnil ($B_1 = 1,354$ při stejné vzdálených kategoriích a 1,370 při nestejně vzdálených kategoriích), kvadratický efekt zůstal stejný ($B_2 = -0,168$ v obou případech), kubický efekt se výrazně zmenšil ($B_3 = -0,208$ při stejné vzdálených kategoriích a při nestejně vzdálených kategoriích $B_3 = -0,014$).

Uvažujme opět případ čistého lineárního vztahu mezi logity, tentokrát ale za předpokladu nestejně vzdálených kategorií s metrikou (1, 2, 4, 5). Předpokládejme, že místo hodnot (-0,990), (-0,402), 0,482, 0,734 máme hodnoty (-0,99), (-0,39), 0,81, 1,41. Body (1; -0,99), (2; -0,39), (3; 0,81), (4; 0,81) neleží na žádné přímce, ale

body (1; -0,99), (2; -0,39), (4; 0,81), (5; 1,41) ano. Je to opět přímka $y = 0,60x - 1,59$. Když pro tyto nové logity spočteme hodnoty B_1, B_2, B_3 , zjistíme, že

$$B_1 = -0,632 (-0,99) - 0,316 (-0,39) + 0,316 (0,81) + 0,632 (1,41) = 1,640, \exp(B_1) = 5,155,$$

$$B_2 = 0,5 (-0,99) - 0,5 (-0,39) - 0,5 (0,81) + 0,5 (1,41) = 0, \exp(B_2) = 1,$$

$$B_3 = -0,316 (-0,99) + 0,632 (-0,39) - 0,632 (0,81) + 0,316 (1,41) = 0, \exp(B_3) = 1.$$

Přítomný je jen lineární efekt všech kategorií, kvadratický a kubický je nulový. Platí i opačné tvrzení: Když pro nějaké logity L_1, L_2, L_3, L_4 jsou B_2 a B_3 rovný nule, pak body (1, L_1), (2, L_2), (4, L_3), (5, L_4) leží na přímce.

Speciální kontrasty

V proceduře Logistic Regression si může uživatel zadat svoje vlastní kontrasty. Má-li kategorizovaná proměnná I kategorií, pak matice kontrastů má $I - 1$ řádků, I sloupců nebo I řádků a I sloupců v případě, že zadáme také řádek s prvky $1/I$ pro výpočet průměru. Předpokládejme matici typu $(I - 1) \times I$, tj. matici s $I - 1$ řádky a I sloupci. Tato matice je matice kontrastů, jestliže součet prvků v každém jejím řádku se rovná nule a jestliže řádky jsou lineárně nezávislé. Představíme-li si řádky matice kontrastů jako vektory $c_1, c_2, \dots, c_{I-1}$, pak tyto vektory jsou lineárně nezávislé, právě když z rovnice $a_1 c_1 + a_2 c_2 + \dots + a_{I-1} c_{I-1} = 0$ plyne, že všechny koeficienty $a_1, a_2, \dots, a_{I-1}$ jsou rovný 0. Lineární nezávislost řádků nemusíme předem ověřovat. Pokud by se stalo, že tato podmínka není splněná, procedura ohlásí lineární závislost a ukončí se. Kontrasty jsou ortogonální, jestliže součet součinů odpovídajících si prvků ve všech dvojicích různých řádků se rovná nule. Kontrasty Helmertovy, Difference a Polynomial jsou ortogonální.

Speciální kontrasty mohou být ortogonální, ale nemusí. Příkladem speciálních kontrastů, které nejsou ortogonální, je porovnání efektu první kategorie s průměrným efektem všech dalších, porovnání průměrného efektu první a druhé kategorie s průměrným efektem všech dalších, porovnání průměrného efektu

Tabulka 24. Matice speciálních kontrastů

1/4	1/4	1/4	1/4
1	-1/3	-1/3	-1/3
1/2	1/2	-1/2	-1/2
1/3	1/3	1/3	-1

Zdroj: autorka.

Tabulka 25. Upravená matice speciálních kontrastů (bez prvního řádku)

3	-1	-1	-1
1	1	-1	-1
1	1	1	-3

Zdroj: autorka.

Tabulka 26. Designová matice k upravené matici speciálních kontrastů

1/4	0	0
-1/4	1/2	0
0	-1/2	1/4
0	0	-1/4

Zdroj: autorka.

Tabulka 27. Odhady parametrů a testy významnosti (speciální kontrasty)

	B	S.E.	Wald	df	Sig.	Exp(B)
Edu			59,786	3	,000	
Edu(1)	-3,784	,829	20,832	1	,000	,023
Edu(2)	-2,609	,354	54,171	1	,000	,074
Edu(3)	-3,112	,576	29,224	1	,000	,045
Constant	-,044	,089	,251	1	,617	,957

Zdroj: Naše společnost 2006.

první, druhé a třetí kategorie s průměrným efektem všech dalších, až nakonec porovnání průměrného efektu prvních $I - 1$ kategorií s efektem poslední kategorie. Pro $I = 4$ má matice těchto kontrastů včetně řádku udávajícího význam konstanty tvar, který je v tabulce 24.

Pomocí zlomků matici kontrastů zadat nelze, neboť symbol / má v jazyce SPSS speciální význam.¹² Je možné ji zadat v podobě desetinných čísel s desetinnou tečkou, mohou ovšem nastat problémy s přesností. Nesmíme také při zaokrouhlování zapomenout, že součet čísel v každém řádku kromě prvního se musí rovnat nule. Je rovněž možné vynásobit každý řádek matice kontrastů takovým číslem, aby všechny prvky nově vzniklé matice byly celá čísla. To potom ale musíme vzít v úvahu při interpretaci parametrů B, jak bude patrné z dalšího textu. Druhý řádek matice z tabulky 24 vynásobíme třemi, třetí dvěma a čtvrtý třemi a využijeme toho, že první řádek zadávat nemusíme. Tak dostaneme speciální matici kontrastů o třech řádcích a čtyřech sloupcích (viz tabulku 25). Designová matice k tabulce 25 a výsledky jsou v tabulkách 26 a 27.

Parametr B_0 je průměrný efekt všech kategorií jako obvykle. Význam parametrů B_1 , B_2 , B_3 je dán typem kontrastů uvedeným v tabulce 25. To znamená, že

¹² To se může změnit, v některých procedurách už je to možné.

B_1 je rozdíl trojnásobku efektu první kategorie a součtu efektů ostatních kategorií, B_2 je rozdíl součtu efektů první a druhé kategorie a součtu efektů třetí a čtvrté kategorie, B_3 je rozdíl součtu efektů prvních tří kategorií a trojnásobku efektu čtvrté kategorie. Význam, který chceme, je v matici kontrastů z tabulky 24 a ten mají parametry $B_1/3$, $B_2/2$, $B_3/3$ (jednotlivé parametry musíme vydělit čísly, kterými jsme násobili odpovídající řádky matice kontrastů, abychom dostali celá čísla).

$$B_1/3 = -0,990 - 1/3(-0,402 + 0,482 + 0,734) = -1,261, \exp(B_1/3) = 0,283,$$

$$B_2/2 = 1/2(-0,990 - 0,402) - 1/2(0,482 + 0,734) = -1,304, \exp(B_2/2) = 0,271,$$

$$B_3/3 = 1/3(-0,990 - 0,402 + 0,482) - 0,734 = -1,037, \exp(B_3/3) = 0,354.$$

Standardní chyby (S.E.) parametrů $B_1/3$, $B_2/2$, $B_3/3$ jsou $0,829/3 = 0,276$, $0,354/2 = 0,177$, $0,576/3 = 0,192$, kde 0,829, 0,354, 0,576 jsou standardní chyby parametrů B_1 , B_2 , B_3 z tabulky 27. Hodnoty Wald, df a Sig. se nemění.

Závěr

V článku bylo zmíněno odděleně několik typů kontrastů, bude proto účelné provést jejich porovnání na menším prostoru. Pokud se nejprve zaměříme na význam konstanty v rovnici logistické regrese, pak se z řady vymykají kontrasty typu Indicator. Pouze u nich má konstanta význam logitu pro referenční kategorii. U všech ostatních se konstanta rovná průměru z logitů pro jednotlivé kategorie vysvětlující proměnné. Vzhledem k typu vysvětlující proměnné lze doporučit následující strategii. Pro nominální proměnné jsou vhodnými typy kontrastů Indicator, Simple nebo Deviation. U kontrastů Indicator nebo Simple zvolíme jednu kategorii za referenční a efekty ostatních kategorií porovnáваме s efektem této referenční kategorie. Referenční kategorie má význam jakési kontrolní skupiny. Mezi kontrasty Indicator a Simple vybíráme podle toho, jaký význam chceme dát konstantě. Pokud žádné z kategorií nechceme dát význam kontrolní skupiny, zvolíme kontrasty Deviation. V tom případě porovnáваме efekty jednotlivých kategorií s průměrným efektem všech kategorií.

Kontrasty Indicator, Simple a Deviation lze použít vždy, tedy i pro proměnné ordinální, pro které lze navíc doporučit kontrasty Repeated a Polynomial. Kontrasty Repeated porovnáваме efekty sousedních kategorií, a mohou tak odhalit existenci trendu v efektech jednotlivých kategorií. Pokud se existence trendu potvrdí, lze zjistit pomocí kontrastů Polynomial, o jaký trend se jedná. Pokud se existence trendu nepotvrdí, je lépe použít kontrasty Indicator, Simple nebo Deviation. Kontrasty Helmertovy a Difference (obrácené Helmertovy kontrasty) jsou vhodné tehdy, když existuje apriorní logické uspořádání kategorií (zde slovo „uspořádání“ může a nemusí mít význam ordinality), jako je tomu na příklad u proměnné s kategoriemi 1 = za prací nejsem ochoten dojíždět mimo

Tabulka 28. Speciální kontrasty pro proměnnou socioekonomické postavení: verze 1

-1/5	-1/5	-1/5	1	-1/5	-1/5
1/3	1/3	1/3	0	-1/2	-1/2
-1/2	-1/2	1	0	0	0
1	-1	0	0	0	0
0	0	0	0	1	-1

Zdroj: autorka.

Tabulka 29. Speciální kontrasty pro proměnnou socioekonomické postavení: verze 2

-1/5	-1/5	-1/5	1	-1/5	-1/5
1	-1	0	0	0	0
0	1	-1	0	0	0
0	0	1	0	-1	0
0	0	0	0	1	-1

Zdroj: autorka.

místo svého bydliště, 2 = jsem ochoten dojíždět, ale jen do vzdálenosti 20 km, 3 = jsem ochoten dojíždět i do větší vzdálenosti, než je 20 km, ale nejvýše do 50 km, 4 = jsem ochoten dojíždět i na větší vzdálenost, než je 50 km, 5 = nevím, ještě jsem o tom nepřemýšlel. Když tuto proměnnou překódujeme $5 \rightarrow 1$, $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 4$, $4 \rightarrow 5$, pak Helmertovy kontrasty porovnávají efekt nerozhodnutých s průměrným efektem rozhodnutých, efekt neochotných s průměrným efektem ochotných, efekt ochotných dojíždět do 20 km s průměrným efektem ochotných dojíždět na větší vzdálenost, efekt ochotných dojíždět do více než 20 km, ale nejvýše do 50 km s efektem ochotných dojíždět i na větší vzdálenost.

S proměnnou socioekonomické postavení s kategoriemi 1 = vyšší odborníci, 2 = nižší odborníci, 3 = rutinní nemanuální pracovníci, 4 = samostatně výdělečně činní, 5 = kvalifikovaní dělníci, 6 = polokvalifikovaní a nekvalifikovaní dělníci se setkáváme prakticky v každém výzkumu. Pokud bychom s ní pracovali v logistické regresi, s největší pravděpodobností bychom použili kontrasty Indicator, resp. Simple nebo Deviation. Myslím, že v některých kontextech by byly zajímavější speciální kontrasty, které by porovnávaly efekt samostatných s průměrným efektem zaměstnanců, průměrný efekt nemanuálních zaměstnanců s průměrným efektem manuálních zaměstnanců, efekt rutinních nemanuálních pracovníků s průměrným efektem vyšších a nižších odborníků, efekt vyšších odborníků s efektem nižších odborníků, efekt kvalifikovaných dělníků s efektem ostatních dělníků. Matice kontrastů bez prvního řádku, ve kterém by se všechny prvky rovnaly 1/6, by měla tvar, který uvádí tabulka 28.

Existují i jiné možnosti, na příklad porovnat efekt samostatných s průměrným efektem zaměstnanců, a pak použít kontrastů Repeated, neboť třídy zaměstnanců jsou považovány za ordinální. Matice kontrastů (opět bez prvního řádku, kde by se všechny prvky rovnaly 1/6) by měla tvar, který uvádí tabulka 29.

Je-li mezi vysvětlujícími proměnnými více kategorizovaných proměnných než jedna, nepřináší to žádné nové problémy, pokud ovšem mezi nimi neuvažujeme interakce. Téma interakcí se ale do této studie už nevešlo. Za zmínku snad ještě stojí, čemu se v případě více kategorizovaných vysvětlujících proměnných rovná konstanta. Předpokládejme, že máme dvě, na příklad Edu se čtyřmi kategoriemi a Sektor se dvěma (1 = soukromý, 2 = jiný). Když pro obě proměnné zvolíme kontrasty Indicator, pro Edu na příklad Indicator(1) a pro Sektor Indicator(2), bude se konstanta B_0 rovnat logit(Volba) pro referenční kategorie, tj. pro základní vzdělání a jiný sektor. Když pro jednu z proměnných zvolíme kontrasty Indicator(refcat) a pro druhou jakýkoliv z dalších zde uvedených kontrastů, bude se konstanta B_0 rovnat průměru logitů v referenční kategorii. Když na příklad zvolíme pro Edu kontrasty Repeated a pro Sektor Indicator(2), konstanta B_0 se bude rovnat průměru logitů při jiném sektoru, tj. 1/4 (logit(Volba) pro základní vzdělání a jiný sektor + ... + logit(Volba) pro pomaturitní vzdělání a jiný sektor). Když pro obě proměnné zvolíme jiné kontrasty než Indicator(ref), bude se konstanta B_0 rovnat průměru ze všech logitů.

BLANKA ŘEHÁKOVÁ je vědeckou pracovnící Sociologického ústavu AV ČR, v.v.i. V současné době je členkou oddělení Hodnotové orientace ve společnosti. Zaměřuje se na studium široce pojatých hodnot. Zabývá se rovněž statistickým modelováním a sociálními sítěmi, výzkumem sociálních a vzdělanostních nerovností, volebního chování a vztahu k životnímu prostředí.

Literatura

- Bock, R. D. 1975. *Multivariate Statistical Methods in Behavioral Research*. New York: Mc Graw-Hill.
- Hosmer, D. W., S. Lemeshow. 1989. *Applied Logistic Regression*. New York: John Wiley and Sons.
- Menard, S. 1995. „Applied Logistic Regression Analysis.“ *Sage University Paper Series on Quantitative Applications in the Social Sciences* 07-106. Thousand Oaks, CA: Sage.
- SPSS Regression Models™ 9.0. User manual. 1999. Chicago, USA: SPSS Inc.



revue PROSTOR

společnost

politika

kultura

umění

Stati, názory, eseje, polemika, rozhovor a poznámky na téma

Výbuch informatické bomby – a co dál?

Úvodník Milana Hanuše *Anděl mého počítače* ■ názor Ivana M. Havla *Co je to za revoluci?* ■ polemický esej Blumfelda S. M. *Walden aneb Zpráva o životě bez médií* ■ rozhovor Paula Virilia a Friedricha Kittlera *O životě ve stínu informatické bomby* ■ stati Petra Saka *Dva scénáře digitalizovaného světa*, Jana Šterna *Média, pravda a svoboda* a Vladimíra Justa *Řeč (novo)řeči* ■ úvaha Jean Hardyové k výročí Roberta Assagioliho ■ komentář Jana Urbana *Věk šamanů* ■ přednáška Johna Keanea *Jedenáct tezí o komunikační nadbytečnosti* ■ poznámky Eduarda P. Martina, Drahomíry Pithartové, Stanislava Komárka, Romany Pavrovské, Radima Seltenreicha ■ ukázka z knihy Richarda Tarnase *Vesmír a psyché* ■ povídka Toniho Florese *Anima mea*

Nové číslo revue PROSTOR 79 (160 Kč) vyjde v září 2008 a můžete je zakoupit ve vybraných knihkupectvích a v redakci časopisu, či objednat v rámci celoročního předplatného (čtyři čísla za 380 Kč) na adrese redakce revue PROSTOR, Politických vězňů 15, 110 00 Praha 1, tel. + fax: 222248052, e-mail: revue.prostor@atlas.cz, web: www.revueprostor.cz