

2 Logit i logit wielomianowy

2.1 Logit i miary dopasowania

Miary dopasowania modelu do danych empirycznych opisane w niniejszym rozdziale mają szersze zastosowanie, do szerszej klasy modeli z dyskretnymi zmiennymi zależnymi, nie tylko do modelu logitowego. Ukazane oszacowania parametrów modelu logitowego służą ilustracji empirycznej omawianych miar i ich własności.

Ważną cechą tych miar jest fakt, iż inaczej jak to jest w przypadku standardowej miary R^2 dla modelu regresji liniowej, przeważnie nie mogą być one wykorzystywane jako element konstrukcji testów diagnostycznych. Celem ich wyliczenia jest wyłącznie zobrazowanie stopnia dopasowania modelu do danych empirycznych.

Jako dane przykładowe zostanie wykorzystana próbka z amerykańskiego badania *Panel Study of Income Dynamics*, opisana w artykule T.Mroz (1987). Dane zawierają informacje o zamężnych kobietach rasy białej w wieku 30-60 lat i ich uczestnictwie w rynku pracy. Zmienna `lfp` określa status na rynku pracy i przyjmuje wartość 1 dla pracującej kobiety.

```
.use http://www.stata-press.com/data/lf2/binlfp2, clear

.describe

.logit lfp k5 k618 age wc hc lwg inc, nolog
* ilorazy względnych szans
.logit, or
* efekty krańcowe
.mfx

.net search fitstat

.fitstat

Measures of Fit for logit of lfp

Log-Lik Intercept Only:    -514.873    Log-Lik Full Model:      -452.633
D(745):                    905.266    LR(7):                   124.480
                           Prob > LR:                0.000
McFadden's R2:            0.121    McFadden's Adj R2:      0.105
Maximum Likelihood R2:    0.152    Cragg & Uhler's R2:     0.204
McKelvey and Zavoina's R2: 0.217    Efron's R2:             0.155
Variance of y*:           4.203    Variance of error:      3.290
Count R2:                  0.693    Adj Count R2:           0.289
```

AIC:	1.223	AIC*n:	921.266
BIC:	-4029.663	BIC' :	-78.112

Oszacowania nieznanymi parametrów modeli dla dyskretnej zmiennej zależnej uzyskuje się zazwyczaj wykorzystując metodę największej wiarygodności. Podstawowymi miarami dopasowania są wartość logarytmu funkcji wiarygodności dla modelu wyłącznie ze stałą ($\ln L_0$) i wartość logarytmu funkcji wiarygodności dla oszacowanego modelu ($\ln L(full)$). Te dwie wartości stanowią podstawę konstrukcji dla wielu innych miar dopasowania.

Odchylenie (*ang. deviance*) jest statystyką porównującą wartość logarytmu funkcji wiarygodności dla oszacowanego modelu z wartością logarytmu funkcji wiarygodności dla modelu, który ma parametr specyficzny dla każdej obserwacji.

$$D = -2 \ln L(full)$$

Statystyka D nie ma rozkładu χ^2 . Współczynniki determinacji są zdefiniowane w sposób standardowy, ale należy pamiętać, że dla dyskretnej zmiennej zależnej wariancja jest ograniczona z góry.

$$R^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} = \frac{var(\hat{y})}{var(\hat{y}) + var(\varepsilon)} = 1 - \left\{ \frac{L_0}{L(full)} \right\}^{\frac{2}{N}}$$

Gdzie N jest liczbą obserwacji. Statystyka R^2 ma zastosowanie wyłącznie w modelu liniowym prawdopodobieństwa

$$adj - R^2 = \left(R^2 - \frac{K}{N-1} \right) \left(\frac{N-1}{N-K-1} \right)$$

gdzie K jest liczbą zmiennych niezależnych w modelu.

Pozostałe statystyki są w literaturze określane wspólną nazwą pseudo R^2 . Pomimo tego, że ta wielkość jest raportowana dla każdego modelu szacowanego z wykorzystaniem metody największej wiarygodności, dla każdego z nich przyjmuje inną formę i posiada inną interpretację.

Statystyka Mc-Fadden R^2 (Likelihood Ratio Index)

$$LRI = 1 - \frac{\ln L(full)}{\ln L_0}$$

porównuje wartość logarytmu funkcji wiarygodności oszacowanego modelu, z wartością logarytmu funkcji wiarygodności modelu ze stałą jako jedyną zmienną objaśniającą. Można pokazać, że im większa liczba obserwacji jest w próbie na podstawie której szacowane są parametry modelu, tym niższą wartość przyjmuje statystyka R^2 Mc Faddena (LRI). Dodatkowo, ten wskaźnik nigdy nie osiąga wartości 1.

Statystyka pokazuje o ile lepiej dopasowany do danych jest wybrany model od naiwnego modelu, w którym jedyną zmienną objaśniającą jest stała. Zatem, w pewnym sensie, pokazuje w jaki sposób wariancja zmiennych niezależnych pozwala wytłumaczyć zróżnicowanie zmiennej zależnej.

Statystyka Mc-Faddena powiela tę samą nieprzyjemną w zastosowaniach własność, co standardowy współczynnik dopasowania R^2 . Jej wartość zwiększa się po dodaniu kolejnej zmiennej do modelu, dlatego można też korzystać z wartości skorygowanej.

$$\text{adj } LRI = 1 - \frac{\ln L(\text{full}) - K^*}{\ln L_0}$$

gdzie K^* oznacza liczbę parametrów wymagających oszacowania.

Wartość statystyki jest korygowana o liczbę zmiennych uwzględnionych w postaci funkcyjnej modelu.

Cameron i Windmeijer (1993) pokazali, że LRI można dekomponować w analogiczny sposób jak statystykę R^2 w modelu regresji. W tym celu należy wykorzystać ℓ_{MAX} oznaczające maksymalną możliwą do uzyskania wartość logarytmu funkcji wiarygodności, tj. wartość przy doskonałym dopasowaniu modelu do danych. Dla modelu z binarną zmienną zależną i ciągłą dystrybuantą wynosi ona 0. Całkowita wariancja $\ell_{max} - \ln L_0$, czyli maksymalny możliwy do uzyskania przyrost wartości funkcji logarytmu wiarygodności, jest dekomponowany na $L(\text{full}) - L_0$ część wariancji wyjaśnioną przez model, oraz $\ell_{max} - \ln L(\text{full})$ część niewyjaśnioną.

Inna statystyka dopasowania, analogiczna do R^2 w modelu regresji liniowej została zaproponowana przez G.S. Maddalę.

Maximum Likelihood R^2

$$ML R^2 = 1 - \left\{ \frac{\ln L_0}{\ln L(\text{full})} \right\}^{\frac{2}{N}}$$

Jej konstrukcja zakłada, że składnik losowy ma rozkład normalny. Posiada ona zbliżone własności do LRI, jednak jej granica górna jest mniejsza od 1.

Statystyka Cragg & Uhler R^2 jest zmodyfikowaną statystyką Maximum Likelihood R^2 Maddali, znormalizowaną do przedziału $[0, 1]$.

$$R_{CU} = \frac{ML R^2}{\max ML R^2} = \frac{1 - \left\{ \frac{L(0)}{L(\text{full})} \right\}^{\frac{2}{N}}}{1 - \left\{ L(0) \right\}^{\frac{2}{N}}}$$

Wartość tej statystyki, podobnie jak wartość R^2 Mc-Faddena zależy od statystycznej istotności oszacowanego modelu (łącznej istotności ocen nieznanymi parametrów).

Dla zmiennej zależnej przyjmującej dwie wartości można wyznaczyć wartość statystyki Efron R^2 . Jest ona uzależniona od wartości współczynnika korelacji pomiędzy rzeczywistą wartością zmiennej zależnej a wartością dopasowaną wynikającą z modelu. Jest ona zdefiniowana następująco

$$\hat{y} = \hat{\pi} = Pr(y = 1 | X)$$

$$R_{Efron}^2 = 1 - \frac{\sum (y_i - \hat{\pi}_i)^2}{\sum (y_i - \bar{y})^2}$$

Dla modeli objaśniających zróżnicowanie zmiennej ukrytej

$$y^* = X\beta + \varepsilon$$

miarą dopasowania jest McKelvey i Zavoina R^2 . Pokazuje ona jaką, conajmniej, część zróżnicowania zmiennej zależnej można byłoby wytłumaczyć za pomocą zmiennych uwzględnionych w modelu, gdyby wartości zmiennej zależnej były obserwowane.

$$R_{MZ}^2 = \frac{var(\hat{y}^*)}{var(y^*)} = \frac{var(\hat{y}^*)}{var(\hat{y}^*) + var(\varepsilon)}$$

gdzie $var(\hat{y}^*) = \beta' var(X) \beta$

Dla modelu z dyskretnymi zmiennymi zależnymi definiuje się również R^2 liczebnościowe (*Count- R^2*). Pokazuje ono jaki procent wartości zmiennej zależnej można prawidłowo przewidzieć posługując się oszacowanymi wartościami parametrów modelu.

$$Count - R^2 = \frac{1}{N} \sum_j n_{jj}$$

Uwaga: w modelu z binarną, np. 0-1, zmienną zależną zgadując można przewidzieć do 50% prawidłowych wyników.

Powyższa miara może być skorygowana. Jeżeli jeden z wyników (wartości zmiennej zależnej) występuje w danych dużo częściej niż pozostałe, to model wskazujący za każdym razem ten wynik będzie, niezależnie od swojej postaci, charakteryzowany przez wysoką wartość statystyki R^2 liczebnościowego. Z tego powodu miarę koryguje się odejmując od liczby prawidłowo przewidzianych wartości liczbę najczęściej pojawiających się wartości zmiennej zależnej

$$AdjCount - R^2 = \frac{\sum_j n_{jj} - \max(n_{r+})}{\max(n_{r+})}$$

Bardziej szczegółowy przegląd miar dopasowania dla modeli z binarną zmienną zależną można znaleźć w Veall, Michael R. i Zimmermann, Klaus F. (1996): *Pseudo- R^2 Measures for Some Common Limited Dependent Variable Models*. Collaborative Research Center 386, Discussion Paper 18

2.2 Porównanie ocen parametrów modeli logitowych i probitowych

Porównanie parametrów pomiędzy różnymi modelami regresji liniowej oszacowanymi na podstawie tej samej próby losowej jest możliwe ponieważ zróżnicowanie zmiennej zależnej jest stałe. W modelu z binarną zmienną zależną stała jest wariancja zmiennej ukrytej y^* , a dodanie zmiennej do modelu powoduje zmianę całkowitej wariancji. Dzieje się tak w efekcie nieobserwowanej „zmiany” zmiennej ukrytej. Prowadzi to do wniosku, iż porównywanie wartości oszacowań współczynników nawet dla zagnieżdżonych modeli może prowadzić do nieprawidłowych wniosków, gdyż wartości oszacowań zależą od skalowania zmiennej zależnej, które może być różne.

Rozważmy przykład, w którym modele różnią się jedną statystycznie nieistotną zmienną objaśniającą.

```
. logit lfp k5 k618 age wc hc lwg inc, nolog
```

```
Logistic regression                Number of obs   =       753
                                LR chi2(7)       =      124.48
                                Prob > chi2        =       0.0000
Log likelihood = -452.63296        Pseudo R2      =       0.1209
```

```
-----+-----
            lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
            k5 | -1.462913   .1970006    -7.43  0.000   -1.849027   -1.076799
           k618 | -.0645707   .0680008    -0.95  0.342   -1.1978499   .0687085
            age | -.0628706   .0127831    -4.92  0.000   -.0879249   -.0378162
             wc |  .8072738   .2299799     3.51  0.000   .3565215    1.258026
             hc |  .1117336   .2060397     0.54  0.588   -.2920969    .515564
            lwg |  .6046931   .1508176     4.01  0.000   .3090961    .9002901
             inc | -.0344464   .0082084    -4.20  0.000   -.0505346   -.0183583
           _cons |  3.18214   .6443751     4.94  0.000   1.919188    4.445092
-----+-----
```

```
. predict p1, xb
```

```
. logit lfp k5 age wc hc lwg inc, nolog
```

```
Logistic regression                Number of obs   =       753
                                LR chi2(6)       =      123.58
                                Prob > chi2        =       0.0000
Log likelihood = -453.08434        Pseudo R2      =       0.1200
```

```
-----+-----
            lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
            k5 | -1.438114   .1939502    -7.41  0.000   -1.818249   -1.057979
            age | -.0577442   .0115086    -5.02  0.000   -.0803007   -.0351877
-----+-----
```

wc		.8183694	.2294614	3.57	0.000	.3686333	1.268105
hc		.1100686	.205615	0.54	0.592	-.2929294	.5130666
lwg		.615903	.1503474	4.10	0.000	.3212275	.9105784
inc		-.0349261	.0081766	-4.27	0.000	-.050952	-.0189002
_cons		2.865892	.5469947	5.24	0.000	1.793802	3.937982

```
-----
```

```
. predict p2,xb
```

```
. gen x = _n
```

```
. sc p1 p2 x
```

```
. sum p1 p2
```

Variable		Obs	Mean	Std. Dev.	Min	Max
p1		753	.3137249	.9554016	-4.255506	3.234711
p2		753	.3139314	.9478656	-4.288973	3.103368

Jak widać na rysunku wartości dopasowane z obu modeli różnią się, co nie jest zaskoczeniem, gdyż inny jest zestaw zmiennych objaśniających. Jednocześnie to co stanowi problem to różną wartość wariancji oraz średnich dla zmiennej ukrytej w obu modelach. Powoduje to, że wartości ocen parametrów są nieporównywalne.

Z tego powodu dla modeli wyborów dyskretnych znacznie wygodniejszym sposobem interpretacji wyników jest wykorzystanie efektów krańcowych.

```
. mfx
```

```
Marginal effects after logit
```

```
y = Pr(lfp) (predict)
```

```
= .57779421
```

variable		dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
k5		-.3568748	.04821	-7.40	0.000	-.451366 - .262383	.237716
k618		-.0157519	.01659	-0.95	0.342	-.048266 .016763	1.35325
age		-.0153371	.00311	-4.93	0.000	-.021434 - .00924	42.5378
wc*		.1880592	.05003	3.76	0.000	.09001 .286109	.281541
hc*		.0271985	.05004	0.54	0.587	-.070882 .125279	.391766
lwg		.1475137	.03674	4.01	0.000	.075496 .219532	1.09711
inc		-.0084031	.002	-4.19	0.000	-.012332 - .004474	20.129

```
-----
```

```
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. mfx
```

```

Marginal effects after logit
  y = Pr(lfp) (predict)
    = .57784458

```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
k5	-.3508138	.04745	-7.39	0.000	-.443814 - .257814	.237716
age	-.0140861	.0028	-5.03	0.000	-.019573 - .008599	42.5378
wc*	.1904694	.04979	3.83	0.000	.092875 .288064	.281541
hc*	.0267932	.04994	0.54	0.592	-.071089 .124675	.391766
lwg	.1502435	.03663	4.10	0.000	.078452 .222035	1.09711
inc	-.0085199	.002	-4.27	0.000	-.012434 -.004606	20.129

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Obliczone oszacowania efektów krańcowych są niezależne od przedziału wartości dla oszacowania zmiennej ukrytej.

2.3 Porównanie logitu i wielomianowego logitu

2.3.1 Wielomianowy model logitowy

Do modelowania zjawiska należy użyć modelu wielomianowego logitowego, gdy zmienna zależna jest dyskretna o skończonej liczbie wartości (niewielkiej), wartości zmiennej nie są uporządkowane, rozkład logistyczny w sposób rozsądny opisuje prawdopodobieństwo wyboru, oraz w bazie obserwacji zawarte są informacje o charakterystykach decydentów.

O modelu wielomianowym możemy myśleć jako o serii jednocześnie szacowanych modeli dla binarnej zmiennej zależnej. Dane wykorzystane w kolejnym przykładzie pochodzą z amerykańskiego badania General Social Survey z 1982 roku. Zmienna *occ3* oznacza zawód respondenta, kodowany przez 3 kategorie: pracownik fizyczny (ang. *manual worker*), pracownik umysłowy (ang. *white collar worker*), specjalista (ang. *professional*). Ale wyniki oszacowań trzech modeli binarnych będą różnić się od wyników modelu wielomianowego z dwóch powodów:

1. Trzy modele dla zmiennej binarnej zawierają zbędną informację
2. Każdy z modeli binarnych jest szacowany na podstawie innej grupy obserwacji

```

/* ML vs L                                     */
/* Long & Freese(2003)                         */
.use http://www.stata-press.com/data/lf2/nomintro2, clear

```

```
.tab prof_man, miss  
  
.logit prof_man ed, nolog  
  
.tab wc_man, miss  
  
.logit wc_man ed, nolog  
  
.tab prof_wc, miss  
  
.logit prof_wc ed
```

Pierwszy z modeli binarnych został oszacowany na podstawie próby liczącej 296 obserwacji, drugi na podstawie próby o liczebności 225 obserwacji, a ostatni na podstawie próby 153 obserwacji, podczas gdy zbiór zawiera 337 obserwacji.

```
.tab occ3, miss  
  
.mlogit occ3 ed, nolog  
  
.mlogit occ3 ed, nolog baseoutcome(1)  
  
.mlogit, rrr  
  
.mlogit occ3 ed, nolog robust  
  
.sum white ed exper  
  
.tab occ, missing  
  
.mlogit occ white ed exper, baseoutcome(5)
```

Ponadto, oszacowania ocen parametrów w modelach dla zmiennych binarnych odbiegają od wartości oszacowań uzyskanych w modelu wielomianowym. Dzieje się tak dlatego, że szacując oddzielnie modele binarne nie jest narzucane ograniczenie o równości współczynników modelu, tak jak się to dzieje w modelu wielomianowym.

2.4 Porównanie wielomianowego i warunkowego logitu

2.4.1 Warunkowy model logitowy

Do modelowania zjawiska należy użyć modelu warunkowego logitowego, gdy zmienna zależna jest dyskretna o skończonej liczbie wartości (niewielkiej), wartości zmiennej nie są uporządkowane, gdy rozkład logistyczny w sposób rozsądny opisuje prawdopodobieństwo wyboru, oraz w bazie obserwacji zawarte są informacje o charakterystykach alternatyw.

```
use http://www.stata-press.com/data/lf2/travel2.dta
```

```
describe
```

Oszacowanie warunkowego modelu logitowego wymaga aby zbiór danych został przygotowany w specyficzny sposób. Dane do przykładu pochodzą z artykułu Greene i Hensher (1997). Dotyczą wyboru środka transportu podczas podróży na wakacje. Dostępne są trzy środki transportu: kolej, autobus i samochód.

```
list id mode train bus time invc choice in(1/9), sepby(id)
```

Zmienna `id` jest identyfikatorem obserwacji. Jak widzimy każdej obserwacji odpowiada trzy wiersze danych. Pierwszy wiersz opisuje charakterystyki podróży pociągiem, drugi charakterystyki podróży autobusem, a trzeci charakterystyki podróży samochodem. Zmienna `choice` koduje informacje o dokonanym wyborze przyjmując wartość 1 dla wybranego środka transportu.

Tym co odróżnia składnię polecenia warunkowego modelu logitowego od innych poleceń w programie Stata, to fakt że niezbędne jest zadeklarowanie identyfikatora grupy obserwacji. W tym celu wykorzystywana jest opcja `group(nazwa)`

```
clogit choice train bus time invc, group(id)
```

Ujemne oszacowania wartości parametrów przy zmiennych `time` oraz `invc` wskazują, że wraz ze wzrostem wartości tych zmiennych mniej prawdopodobny staje się wybór środka transportu, który rzeczywiście został wybrany przez respondenta.

```
clogit ,or
```

```
predict prob
```

```
list train bus time invc choice prob in 1/3
```

Dane wykorzystane w kolejnym przykładzie ponownie pochodzą z amerykańskiego badania General Social Survey z 1982 roku. W trakcie badania pytano respondentów o wykonywany zawód, który klasyfikowano za pomocą pięciu kategorii: prace pomocnicze (ang. *menial jobs*), prace robotnicze (ang. *blue collar jobs*), prace rzemieślnicze (ang. *craft jobs*), prace urzędnicze (ang. *white collar jobs*), oraz prace eksperckie (ang. *professional jobs*). W przykładzie ograniczono się do trzech zmiennych niezależnych: `white`, która jest indykatorem respondenta rasy białej, `ed` oznaczającej liczbę lat spędzonych w systemie edukacyjnym oraz `exper` oznaczającej doświadczenie zawodowe mierzone w latach pracy.

Ponieważ model wielomianowy i model warunkowy są matematycznie równoważne to oszacowania parametrów modelu wielomianowego można uzyskać wykorzystując komendy do szacowania komendy do szacowania modelu warunkowego. W tym celu należy rozszerzyć zbiór danych, oraz przekształcić zmienne objaśniające w interakcje.

Wartością dodaną takiej operacji jest fakt, iż pozwala ona na szacowanie modeli mieszanych, w których część informacji jest gromadzona na poziomie decydentów, a część opisuje dostępne alternatywy. Dodatkowo, można wprowadzić pewne ograniczenia na parametry, które nie są możliwe do uzyskania w standardowym modelu wielomianowym.

W pierwszym kroku szacujemy parametry modelu wielomianowego.

```

/*****/
/* MNL jako CL                               */

use http://www.stata-press.com/data/lf2/nomocc2.dta, clear

mlogit occ white ed exper

```

Polecenie `gen id=_n` tworzy unikalny identyfikator dla obserwacji którego wartość odpowiada numerowi wiersza macierzy obserwacji. Do oszacowania modelu warunkowego potrzebujemy 5 obserwacji dla każdego decydenta, ponieważ jest pięć możliwości wyboru zawodu. Aby je uzyskać wykorzystujemy polecenie `expand`. Multiplikuje ono każdą obserwację.

```

gen id= _n

expand 5

```

W kolejnym kroku sortujemy zbiór względem identyfikatora danych, a następnie tworzymy zmienną `alt`, której wartość wskazuje numer wybranej alternatywy.

```
sort id

gen alt = mod(_n, 5)

replace alt = 5 if alt == 0
```

W następnym kroku stworzymy cztery zmienne zero-jedynkowe wskazujące czy respondent wykonuje dany zawód. Zmienne te następnie wykorzystujemy do stworzenia zmiennej `choice` oznaczającej dokonany wybór

```
gen menial = (alt==1)

gen bluecol = (alt==2)

gen craft = (alt==3)

gen whitecol = (alt==4)

gen choice = (occ==alt)
```

W tym momencie zbiór danych wygląda następująco.

```
list id menial bluecol craft whitecol in 1/10, sepby(id)
```

Ostatnim krokiem jest utworzenie interakcji przez pomnożenie czterech zmiennych wskazujących poszczególne wykonywane zawody przez zmienne niezależne. W ten sposób utworzyliśmy zmienne objaśniające, których wartości są specyficzne dla konkretnych alternatyw.

```
gen whiteXm = white*menial
gen whiteXbc = white*bluecol
gen whiteXc = white*craft
gen whiteXwc = white*whitecol

gen edXm = ed*menial
gen edXbc = ed*bluecol
gen edXc = ed*craft
gen edXwc = ed*whitecol

gen experXm = exper*menial
gen experXbc = exper*bluecol
gen experXc = exper*craft
```

```
gen experXwc = exper*whitecol
```

```
list id menial bluecol craft whitecol edXm edXbc edXc edXwc in 1/5
```

Ostatnim krokiem jest wykorzystanie utworzonych interakcji jako zmiennych niezależnych przy szacowaniu warunkowego modelu logitowego.

```
cllogit choice whiteXm edXm experXm menial whiteXbc edXbc experXbc /*  
*/ bluecol whiteXc edXc experXc craft whiteXwc edXwc experXwc /*  
*/ whitecol, group(id)
```

Uzyskane wartości oszacowań parametrów są identyczne jak w modelu wielomianowym. Również ich interpretacja jest taka sama.

Literatura

- [1] Cameron, A.C. i Windmeijer, F.A.G. (1993): *R-Squared Measures for Count Data Regression Models with Applications to Health Care Utilization*, Dept. of Economics Working Paper 93-24, University of California at Davis.
- [2] Long, J.S. i Freese J. (2003) *Regression Models for Categorical Dependent Variables Using Stata*, Stata Press, wydanie 2.
- [3] Mroz T. (1987) *The Sensivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions*, *Econometrica*, vol 55(4), pp. 765-799.
- [4] Veall, Michael R. i Zimmermann, Klaus F. (1996) *Pseudo-R² Measures for Some Common Limited Dependent Variable Models*. Collaborative Research Center 386, Discussion Paper 18.
- [5] Williams Richard (2011) *Comparing Logit and Probit Coefficients Between Models and Across Groups*.