# 1 Metoda Największej Wiarogodności

#### 1.0.1 Wprowadzenie

Dotychczas omawiane modele szacowane były za pomocą metody najmniejszych kwadratów. Jednak takie modele nie wyczerpują wszystkich możliwości. Istnieje duża rodzina modeli których nie da się oszacować za pomocą tej metody. W takich przypadkach stosuje się metodą największej wiarogodności.

## 1.1 MNW

Metoda największej wiarogodności polega takim wyborze wartości dla szacowanych parametrów, że maksymalizują one funkcję wiarogodności.

**Lemat 1** *Wiarogodność jest to funkcja $L : \Theta \to \mathbb{R}$ dana wzorem:*

$$L(\theta) = f(\theta, x_1, \ldots, x_n)$$

Funkcja wiarogodności jest to funkcja gęstości prawdopodobieństwa, ale rozważana jako funkcja parametru lub zbioru parametrów $\theta$ przy ustalonych wartościach z próby $x_1, \ldots, x_n$.

**Lemat 2** *$\hat{\theta}$ jest estymatorem największej wiarogodności nieznanego parametru $\theta$ jeżeli:*

$$f(\hat{\theta}(x_1, \ldots, x_n); x_1, \ldots, x_n) = \sup_{\theta \in \Theta} f(\theta, x_1, \ldots, x_n)$$

## 1.2 Modele Wyborów Dyskretnych

Wiele zjawisk będących obiektem badania w naukach społecznych ma charakter dyskretny. Dyskretne zmienne zależne powstają w wyniku wyboru jednej z wieku możliwych alternatyw. Liczba potencjalnych wyborów jest skończona. Modele oparte na zasadzie regresji liniowej nie opisują ich z dużą dokładnością, bowiem zmienna zależna w takim przypadku nie jest ciągła. Przykładami zjawisk ekonomicznych o charakterze dyskretnym są np. podjęcie decyzji o zakupie dobra trwałego przez gospodarstwo domowe, czy jego posiadanie, status aktywności zawodowej, czyli to czy osoba pracuje czy jest bezrobotna, czy fakt, że osoba się uczy bądź nie. Te zjawiska łączy wspólna cecha: dla każdego z nich możliwe są dwa stany zmiennej, będącej przedmiotem zainteresowania. Jednak nie wyczerpuje to wszystkich możliwości zastosowania modeli wyborów dyskretnych. Jeżeli te same zjawiska będziemy badali od innej strony, np. nie fakt czy osoba pracuje czy nie, ale jaki ma zawód, nie fakt czy osoba się uczy czy nie, ale jakie osiągnęła najwyższe wykształcenie, to w takim przypadku każde zjawisko ma więcej niż dwa stany. Takie zjawiska również można analizować za pomocą modeli wyborów dyskretnych.

### 1.2.1 Model losowej użyteczności

Podstawą teoretyczną dla modeli wyborów dyskretnych jest model losowej użyteczności (*Random Utility Model*). W tym modelu jednostka $i$ dokonująca wyboru może go dokonać spośród $J$ dostępnych alternatyw. Z każdą z nich związany jest pewien poziom osiąganej użyteczności. Uzyskiwana użyteczność z wyboru możliwości $j$ wynosi:

$$U_{i,j} = X_{i,j} + \varepsilon_{i,j} \tag{1}$$

Można ją rozłożyć na dwie części. Zależną od ujawnionych parametrów związanych ze zmiennymi $X_{i,j}$, oraz nieujawnionych $\varepsilon_{i,j}$. O części ujawnionej zakładamy, że jest wszystkim znana. W jej skład zazwyczaj wchodzą różne charakterystyki społeczno-demograficzne oraz czynniki związane z przedmiotem wyboru. Część nieujawniona jest znana jedynie podejmującemu decyzję, natomiast badacz tworząc model nie zna jej i traktuje ją jako zmienną losową o pewnym rozkładzie prawdopodobieństwa. Zakładamy, że decydent postępuje racjonalnie i dokonuje wyboru maksymalizując swoją użyteczność. Budujemy model, w którym prawdopodobieństwo takiego wyboru wynosi:

$$Pr(y_j = j) = Pr(U_j > U_k \quad \forall j \neq k)$$

Analityczna postać modelu zależy od wyboru rozkładu zaburzeń losowych.

### 1.2.2 Logit i probit

Rozpoczniemy od najprostszego przypadku, w którym zmienna zależna ma charakter dyskretny i przyjmuje tylko dwie wartości. Celem jest oszacowanie parametrów modelu, w którym możliwe są dwa stany: zjawisko nie wystąpiło, zazwyczaj oznaczane przez 0, oraz stan w którym zjawisko wystąpiło, oznaczane przez 1, to model (1) możemy zapisać jako:

$$Pr(y_i = 1) = F(\theta, X_i) \tag{2}$$

czyli prawdopodobieństwo wyboru konkretnej alternatywy jest funkcją od obserwowanych charakterystyk $X_i$ i zbioru parametrów $\theta$. Do lat siedemdziesiątych ubiegłego wieku powszechnie przyjmowano za $F(\theta, X_i)$ funkcję liniową, a w konsekwencji liniowy model prawdopodobieństwa. Posiada on następujące właściwości:

$$Pr(y_i = 1) = X'\beta$$
$$Pr(y_i = 0) = 1 - X'\beta$$

wektor parametrów $\beta$ odzwierciedla wpływ poszczególnych zmiennych na prawdopodobieństwo wystąpienia jednej z dwóch możliwych sytuacji. Jego zaletą jest to, że może być szacowany metodą MNK.

$$y_i = X_i'\beta + \varepsilon \tag{3}$$

Ale ma też poważne wady. Po pierwsze postać (3) nie gwarantuje, że wartość dopasowana modelu $x'\beta$, która jest interpretowana jako prawdopodobieństwo wystąpienia stanu zakodowanego jako 1, leży pomiędzy zerem a jedynką. Po drugie, ponieważ zmienna objaśniana przyjmuje tylko dwie wartości 0 i 1, to składnik losowy ma rozkład dwumianowy i jest równy $\varepsilon = -X'\beta$ lub $\varepsilon = 1 - X'\beta$. Wobec tego można pokazać, że warunkowa wariancja wynosi:

$$var(\varepsilon \mid X) = X'\beta(1 - X'\beta)$$

Ponieważ $X'\beta$ może przyjmować różne wartości, nie możemy mieć pewności że wariancja w modelu liniowym prawdopodobieństwa będzie nieujemna. Ponadto, składnik losowy jest heteroscedastyczny.

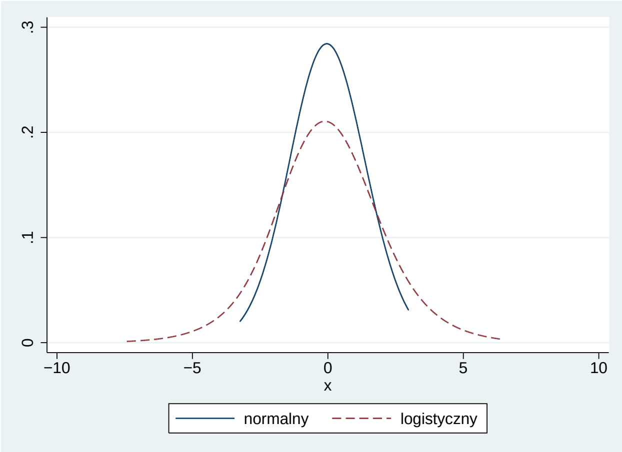
Wobec tych niepożądanych cech modelu liniowego w praktyce ekonometrycznej używa się dwóch innych funkcji prawdopodobieństwa. Przyjęcie rozkładu normalnego prowadzi nas do modelu probitowego:

$$Pr(y_i = 1) = \int_{-\infty}^{X'\beta} \phi(t)dt = \Phi(X'\beta) \tag{4}$$

A przyjęcie rozkładu logistycznego do modelu logitowego.

$$Pr(y_i = 1) = \frac{e^{X'\beta}}{1 + e^{X'\beta}} = \Lambda(X'\beta) \tag{5}$$

Ponieważ obie funkcję są funkcjami prawdopodobieństwa ich wartości są ograniczone przez 0 i 1. Oba rozkłady są symetryczne względem zera. Tym co je różni jest wariancja. Dla rozkładu normalnego wynosi ona 1, dla logistycznego $\frac{\pi^2}{3}$. Oba rozkłady są do siebie bardzo podobne, z tą różnicą że logistyczny ma grubsze ogony. Jeżeli nie dysponujemy duża próbą niemożliwością jest rozróżnić z którego z nich została ona wylosowana. Wobec tego, w modelach jednorównaniowych nie ma znaczenia, którą funkcję wybierzemy, z wyjątkiem sytuacji gdy prawdopodobieństwo skupia się w ogonach rozkładu.




### 1.2.3 Estymacja

Z wyjątkiem modelu prawdopodobieństwa liniowego modele wyborów dyskretnych szacowane są metodą największej wiarogodności. W przypadku modeli dwumianowych każda obserwacja jest traktowana jako realizacja zmiennej losowej z rozkładu Bernoulliego o nieznanych parametrach z prawdopodobieństwem sukcesu $F(x'\beta)$. Łączną funkcję prawdopodobieństwa zapisujemy jako:

$$Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid X) = \Pi_{y_i=0}\Big[1 - F(x'\beta)\Big]\Pi_{y_i=1}\Big[F(x'\beta)\Big]$$

Funkcja wiarogodności dla $n$ obserwacji może być zapisana jako:

$$L(\beta \mid X) = \Pi_{i=1}^{n}\Big[1 - F(x_i'\beta)\Big]^{1-y_i}\Big[F(x_i'\beta)\Big]^{y_i} \tag{6}$$

logarytmując (6) otrzymujemy:

$$\ln L = \sum_{i=1}^{n}\left((1-y_i)\Big[1 - F(x_i'\beta)\Big] + y_i\Big[F(x_i'\beta)\Big]\right) \tag{7}$$

obliczając pochodną funkcji wiarogodności otrzymujemy układ równań normalnych

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n}\left[y_i\frac{f_i}{F_i} + (1-y_i)\frac{-f_i}{1-F_i}\right]x_i = 0 \tag{8}$$

gdzie $F_i$ to dystrybuanta, a $f_i$ funkcja gęstości prawdopodobieństwa. Rozwiązaniem tego układu równań jest wektor parametrów, który nazywamy estymatorem największej wiarogodności i oznaczamy $ENW(\beta)$.

### 1.2.4 Przykład 1.

Przykład ten pochodzi artykułu Thomasa Mroz'a (1987) dotyczącego aktywności kobiet na rynku pracy. Dane do artykułu pochodzą z roku 1976.[1] Zmienna zależna, przynależność do siły roboczej, przyjmuje wartość 1 gdy kobieta jest zatrudniona, 0 w przeciwnym przypadku.

```
. use "binlfp2.dta", clear
 (Data from 1976 PSID-TMroz)
. des lfp k5 k618 age wc hc lwg inc

              storage  display value
variable name   type   format  label      variable label
-----------------------------------------------------------------------------
lfp             byte   %9.0g   lfplbl     Paid Labor Force: 1=yes 0=no
k5              byte   %9.0g              # kids < 6
k618            byte   %9.0g              # kids 6-18
age             byte   %9.0g              Wife's age in years
wc              byte   %9.0g   collbl     Wife College: 1=yes 0=no
hc              byte   %9.0g   collbl     Husband College: 1=yes 0=no
lwg             float  %9.0g              Log of wife's estimated wages
inc             float  %9.0g              Family income excluding wife's

. probit lfp k5 k618 age wc hc lwg inc

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood =  -453.92167
Iteration 2:   log likelihood =  -452.69643
Iteration 3:   log likelihood =  -452.69496

Probit estimates                                  Number of obs   =        753
                                                  LR chi2(7)      =     124.36
                                                  Prob > chi2     =     0.0000
Log likelihood = -452.69496                       Pseudo R2       =     0.1208

------------------------------------------------------------------------------
     lfp |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
      k5 | -.8747112    .1135583    -7.70   0.000    -1.097281   -.6521411
    k618 | -.0385945    .0404893    -0.95   0.340     -.117952    .0407631
     age | -.0378235    .0076093    -4.97   0.000    -.0527375   -.0229095
      wc |  .4883144    .1354873     3.60   0.000     .2227642    .7538645
      hc |  .0571704    .1240052     0.46   0.645    -.1858754    .3002161
     lwg |  .3656287    .0877792     4.17   0.000     .1935847    .5376727
     inc |  -.020525    .0047769    -4.30   0.000    -.0298875   -.0111626
   _cons |  1.918422    .3806536     5.04   0.000     1.172355     2.66449
```

[1]Dane mogą być ściągnięte do Staty poleceniem net from http://www.stata-press.com/data/lfr/, a następnie net install rmcdvs, oraz net get rmcdvs

```
------------------------------------------------------------------------------
```

Interpretacja parametrów modelu probitowego (logitowego także) jest trudna. W zasadzie możemy zinterpretować jedynie znaki przy zmiennych. Znak dodatni oznacza, że większa wartość zmiennej niezależnej zwiększa prawdopodobieństwo przyjęcia przez zmienną zależną wartości 1, a ujemny oznacza, że zwiększenie wartości danej zmiennej objaśniającej zwiększa prawdopodobieństwo otrzymania zera.

W celu interpretacji ekonomicznej zjawiska często oblicza się przyrosty krańcowe (efekty krańcowe). Otrzymuje je się ze wzoru:

$$\frac{\partial Pr(Y=1 \mid X)}{\partial x_k}$$

Pakiet Stata standardowo oblicza krańcowe efekty przyjmując przeciętne wartości pozostałych zmiennych. Nie zawsze są one interesujące i w zastosowaniach warto kontrolować punkt w którym obliczana jest pochodna. Należy pamiętać, że wartość krańcowego efektu zależy od przyjętego poziomu pozostałych zmiennych objaśniających. Dodatkowo, w ogólnym przypadku są one funkcjami nieliniowymi zmiennych objaśniających. Obie cechy wartości krańcowych należy uwzględnić przy interpretacji.

W przypadku objaśniających zmiennych dyskretnych przyrosty krańcowe nie mają sensu. Oblicza się zmianę prawdopodobieństwa, że zmienna $Y$ przyjmie wartość 1, pod wpływem zmiany wartości zmiennej objaśniającej $x_k$ z 0 na 1, według wzoru:

$$\frac{\Delta Pr(Y=1 \mid X)}{\Delta x_k}$$

Wartość zmian dyskretnych prawdopodobieństwa zależy od poziomu startowego zmiennej objaśniającej, jej poziomu końcowego (w przypadku zmiennych zero-jedynkowych to nie ma znaczenia), oraz poziomu pozostałych zmiennych.

Obliczmy efekt krańcowy dla kobiety która ukończyła college i obecnie ma 40 lat.

```
. mfx compute, at(wc=1 age=40)

warning: no value assigned in at() for variables k5 k618 hc lwg inc;
   means used for k5 k618 hc lwg inc

Marginal effects after probit
      y  = Pr(lfp) (predict)
         =  .74021954
```

```
---------------------------------------------------------------------------
variable |      dy/dx  Std. Err.     z     P>|z|  [    95% C.I.   ]      X
---------+-----------------------------------------------------------------
      k5 | -.2836023   .03937   -7.20   0.000  -.360774  -.20643   .237716
    k618 | -.0125133   .01312   -0.95   0.340  -.038238   .013211   1.35325
     age | -.0122633   .00244   -5.03   0.000   -.01704 -.007486        40
     wc*|   .1783511   .04656    3.83   0.000   .087101   .269601         1
     hc*|   .0184605   .04045    0.46   0.648  -.060819    .09774   .391766
     lwg |  .1185456   .03077    3.85   0.000   .058239   .178852   1.09711
     inc | -.0066547   .00158   -4.21   0.000  -.009749  -.00356    20.129
----------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Dla zmiennych, które zadeklarowaliśmy w opcji `at` pochodna została policzona w zadanym punkcie. Dla pozostałych zmiennych Stata automatycznie przyjęła wartości przeciętne. Jak widzimy, dla dwóch zmiennych obliczone zostały zmiany dyskretne, z powodu ich dyskretnego charakteru (oznaczono je na wydruku gwiazdkami przy nazwie zmiennej).

W tym miejscu należy się zastanowić czy efekt dla zmiennej oznaczającej ukończenie przez męża collegu ma sensowną interpretację ekonomiczną. Uzyskany wynik można zinterpretować następująco: *Dla przeciętnej kobiety fakt, że mąż ukończył college zwiększa prawdopodobieństwo uczestnictwa w rynku pracy o 2 %, zakładając że ma ona męża o przeciętnym wykształceniu.* Oczywiście taka interpretacja nie ma sensu ekonomicznego, bowiem nie istnieje "mąż o przeciętnym wykształceniu", ponieważ jest to osoba która w 39 % ukończyła college i w 61 % go nie ukończyła.

Jest to obserwacja, którą można rozszerzyć na wszystkie zmienne o charakterze nominalnym. Dla takich zmiennych efekty krańcowe powinny być liczone dla ich ustalonych wartości, co więcej takich które posiadają sensowną interpretację ekonomiczną.

Sprawdźmy efekty krańcowe dla `hc=0` oraz `hc==1`

```
. mfx compute, at(wc=1 age=40 hc=0)

warning: no value assigned in at() for variables k5 k618 lwg inc;
   means used for k5 k618 lwg inc

Marginal effects after probit
      y  = Pr(lfp) (predict)
         =  .73290575
---------------------------------------------------------------------------
variable |      dy/dx  Std. Err.     z     P>|z|  [    95% C.I.   ]      X
---------+-----------------------------------------------------------------
      k5 | -.2876506    .04272   -6.73   0.000  -.371388 -.203913   .237716
    k618 | -.0126919    .01333   -0.95   0.341   -.03882  .013436   1.35325
     age | -.0124383    .00251   -4.96   0.000  -.017349 -.007527        40
```

```
     wc*|  .1798796    .04573    3.93   0.000   .090258  .269502         1
     hc*|  .0184605    .04045    0.46   0.648  -.060819   .09774         0
    lwg |  .1202378    .03165    3.80   0.000   .058203  .182273   1.09711
    inc | -.0067497    .00169   -3.99   0.000  -.010063 -.003436    20.129
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx compute, at(wc=1 age=40 hc=1)

warning: no value assigned in at() for variables k5 k618 lwg inc;
   means used for k5 k618 lwg inc

Marginal effects after probit
      y  = Pr(lfp) (predict)
         =  .75136622
------------------------------------------------------------------------------
variable |      dy/dx  Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      k5 | -.2771541    .03763   -7.36   0.000  -.350915 -.203393   .237716
    k618 | -.0122287    .01282   -0.95   0.340  -.037353  .012895   1.35325
     age | -.0119845    .00244   -4.91   0.000  -.016773 -.007196        40
     wc*|  .1758324     .0485    3.63   0.000    .08078  .270885         1
     hc*|  .0184605    .04045    0.46   0.648  -.060819   .09774         1
    lwg |  .1158502     .0302    3.84   0.000    .05666  .175041   1.09711
    inc | -.0065034    .00144   -4.51   0.000  -.009333 -.003674    20.129
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Jak widać dla wielkości efektu krańcowego zmiennej oznaczającej wykształcenie męża `hc` punkt w którym jest liczona pochodna nie ma znaczenia. Natomiast wartości oszacowań przy pozostałych zmiennych różnią się, co świadczy o nieliniowym wpływie pozostałych czynników na prawdopodobieństwo posiadania pracy przez kobietę.

### 1.2.5 Miary dopasowania modelu

Zarówno model probit, jak również model logit nie są liniowe względem swoich parametrów. Z tego powodu tradycyjne miary dopasowania modelu do danych empirycznych nie mają zastosowania. Standardową statystyką podawaną w tych modelach jest logarytm ilorazu wiarogodności. Na jego podstawie są tworzona jest większość miar.

Miarą dopasowania zastępującą $R^2$ jest statystyka oparta na ilorazie funkcji wiarogodności. W literaturze ekonometrycznej występuje pod kilkoma nazwami, najbardziej popularną jest iloraz wiarogodności (*Likelihood Ratio Index (LRI)*). Inne nazwy tej statystyki to $pseudo - R^2$, lub $McFadden - R^2$. Mierzy ona łączną istotność oszacowanych parametrów modelu, w stosunku do hipotezy zerowej, że wszystkie parametry modelu są równe zero. Wartość statystyki jest obliczona na postawie wartości funkcji logarytmu wiarogodności. LRI jest zdefiniowany następująco:

$$\rho = 1 - \frac{L(\beta)}{L_0}$$

gdzie: $L(\beta)$ jest wartością funkcji logarytmu wiarogodności dla wyestymowanych parametrów, a $L_0$ jest wartością tej funkcji, gdy wszystkie parametry są równe zero, czyli gdy jedynym regresorem jest stała. Pomimo tego że statystyka LRI przyjmuje wartości z przedziału (0,1), nie ma ona takiej interpretacji jak statystyka $R^2$ dla modelu regresji liniowej. Nawet gdy model jest idealnie dopasowany do danych, statystyka LRI przyjmuje wartości znacznie mniejsze od 1. Wobec tego nie jest ona dobrą miarą dopasowania modelu. Na jej podstawie możemy jedynie stwierdzić który z wyestymowanych modeli na tej samej próbie danych, z tymi samymi zmiennymi objaśniającymi, jest lepiej do nich dopasowany. Natomiast test nic nie mówi o stopniu dopasowania.

Podobną miarą jest Maximum Likelihood $R^2$. Wyraża się ona wzorem:

$$MLR^2 = 1 - \frac{1}{1 + 2\frac{L(\beta)-L_0}{n}}$$

powiela ona zalety oraz wady statystyki LRI.

```
. fitstat

Measures of Fit for probit of lfp

Log-Lik Intercept Only:    -514.873   Log-Lik Full Model: -452.695
D(745):                     905.390   LR(7):                124.356
                                      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:    0.247   Efron's R2:             0.154
Variance of y*:               1.328   Variance of error:      1.000
Count R2:                     0.687   Adj Count R2:           0.274
AIC:                          1.224   AIC*n:                921.390
BIC:                       4029.539   BIC':                 -77.988
```

Polecenie `fitstat` z dodatku SPost do pakietu Stata podaje również kilka kolejnych modyfikacji współczynnika $R^2$, oraz dodatkowo standardowe kryteria informacyjne Akaike (AIC), oraz Schwarza (BIC).

Rozpatrywany przez nas model możemy również przedstawić w formie logitowej.

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

Iteration 0:   log likelihood =  -514.8732
Iteration 4:   log likelihood = -452.63296

Logit estimates                                   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    -.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
------------------------------------------------------------------------------
```

Współczynniki tego modelu nie mają bezpośredniej interpretacji ale możemy je przedstawić w postaci ilorazów względnych szans.

### 1.2.6 Ilorazy względnych szans

Współczynniki modelu zawierają informacje o wpływie zmiennej na poziom prawdopodobieństwa wystąpienia danej kategorii zmiennej objaśnianej. Ich interpretacja jest trudna. Dlatego zazwyczaj nie dokonuje się jej bezpośrednio. Wylicza się ilorazy szans, których interpretacja jest łatwiejsza. Jest to miara podobna do prawdopodobieństwa. Prawdopodobieństwo mierzy szansę zaistnienia jakiegoś stanu, np. szansa wyrzucenia 6 prawidłową sześcienną kostką do gry wynosi $\frac{1}{6}$. Natomiast rzucając kostką mamy szansę 1 do 5, że otrzymanym wynikiem rzutu będzie 6. Iloraz szans jest to stosunek prawdopodobieństw wystąpienia dwóch zdarzeń przy danym wektorze $x_i$ zmiennych objaśniających.

$$\Omega = \frac{Pr(Y=1)}{Pr(Y=0)} = \frac{Pr(Y=1)}{1-Pr(Y=1)}$$

W modelu logitowym logarytm ilorazu szans jest funkcją liniową parametrów.

$$ln\left\{\frac{Pr(Y=1)}{1-Pr(Y=1)}\right\} = ln\Omega(x) = X'\beta$$

Dodatkowym czynnikiem ułatwiającym interpretację tego współczynnika jest jego konstrukcja. Jest on obliczony bez konieczności odwoływania się do innych kategorii zmiennej objaśnianej. Ilorazy szans interpretujemy następująco: jednostkowa zmiana zmiennej $x_k$ powoduje zmianę szans o $\beta x_k$.

Wyniki estymacji przedstawione za pomocą ilorazów szans wyglądają następująco:

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

Iteration 0:   log likelihood =  -514.8732
Iteration 4:   log likelihood = -452.63296

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

------------------------------------------------------------------------------
     lfp | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
      k5 |  .2315607    .0456176    -7.43   0.000     .1573902    .3406843
    k618 |  .9374698    .0637487    -0.95   0.342      .820493    1.071124
     age |   .939065    .0120042    -4.92   0.000     .9158296    .9628899
      wc |  2.241788    .5155662     3.51   0.000     1.428352    3.518469
      hc |  1.118215    .2303967     0.54   0.588     .7466962    1.674583
     lwg |   1.83069    .2761003     4.01   0.000     1.362193    2.460317
     inc |  .9661401    .0079304    -4.20   0.000     .9507211    .9818092
------------------------------------------------------------------------------
```

Można zauważyć, że przy zmiennych ze znakiem ujemnym pojawiły się współczynniki szans mniejsze od 1, a przy zmiennych ze znakiem dodatnim większe od 1. Ich interpretacja jest następująca. Każde kolejne dziecko (k5) zmniejsza szansę zatrudnienie o $0,23$ przy innych czynnikach stałych. Jeśli kobieta ukończyła college (wc=1), to szansa że pracuje wynosi $2,24$ czyli prawdopodobieństwo, że pracuje wynosi 69%.($Pr(pracuje) = \frac{2{,}24}{3{,}24} = 0,69$) przy innych czynnikach stałych.

Mając oszacowane dwa modele możemy porównać ich współczynniki.

```
. qui probit lfp k5 k618 age wc hc lwg inc
. estimates store probit
. qui logit lfp k5 k618 age wc hc lwg inc
. estimates store logit
. estimates table probit logit

----------------------------------------
    Variable |   probit       logit
-------------+--------------------------
```

```
         k5 | -.87471118    -1.462913
       k618 | -.03859449   -.06457068
        age |  -.0378235   -.06287055
         wc |  .48831439    .80727378
         hc |  .05717035    .11173357
        lwg |  .36562871    .60469312
        inc | -.02052503   -.03444643
      _cons |  1.9184223    3.1821405
----------------------------------------
```

Wyraźnie widać, że przeciętnie współczynniki modelu logitowego są o 1,6 razy większe od odpowiadających im współczynników modelu probitowego. Nie jest to przypadek, tylko reguła, która obowiązuje dla prawdopodobieństw spoza ogonów rozkładu.

### 1.2.7 Test ilorazu wiarogodności

Do porównywania różnych form funkcyjnych modeli możemy posłużyć się testem ilorazu wiarogodności. Jest on zbudowany na podobnej zasadzie do testu LRI. Jest on również oparty na różnicy logarytmów wartości funkcji wiarogodności dla dwóch modeli. Ponieważ logarytm funkcji wiarogodności jest zawsze ujemny, statystyka testowa jest zawsze dodatnia.

$$LR = -2(L_1 - L_0) \sim \chi^2(p(c-1)) \tag{9}$$

gdzie: $L_1$ jest logarytm funkcji wiarogodności modelu z narzuconymi ograniczeniami, $\mathrm{L}_0$ - logarytm funkcji wiarogodności modelu bez ograniczeń, $p$ - ilością zmiennych niezależnych w modelu, $c$ - ilością kategorii zmiennej objaśnianej, co ma znaczenie przy modelach wielomianowych.

Test ten może służyć zarówno do porównywania dwóch modeli, jak i do testowania narzuconych ograniczeń na parametry modelu. Jego zaletą jest to, że pozwala bezpośrednio porównać dwa wyestymowane na tej samej próbie modele.

Spojrzmy na wyniki testu dla naszego modelu.

```
. qui probit lfp k5 k618 age wc hc lwg inc
. qui fitstat, save
. qui logit lfp k5 k618 age wc hc lwg inc
. fitstat, dif force

Measures of Fit for logit of lfp

Warning: Current model estimated by logit, but saved model
estimated by probit
```

```
                         Current          Saved      Difference
Model:                    logit           probit
N:                         753             753          0
Log-Lik Intercept Only:  -514.873        -514.873      0.000
Log-Lik Full Model:      -452.633        -452.695      0.062
D:                        905.266(745)    905.390(745) 0.124(0)
LR:                       124.480(7)      124.356(7)   0.124(0)
Prob > LR:                  0.000           0.000        .
McFadden's R2:              0.121           0.121      0.000
McFadden's Adj R2:          0.105           0.105      0.000
Maximum Likelihood R2:      0.152           0.152      0.000
Cragg & Uhler's R2:         0.204           0.204      0.000
McKelvey and Zavoina's R2:  0.217           0.247     -0.030
Efron's R2:                 0.155           0.154      0.001
Variance of y*:             4.203           1.328      2.875
Variance of error:          3.290           1.000      2.290
Count R2:                   0.693           0.687      0.007
Adj Count R2:               0.289           0.274      0.015
AIC:                        1.223           1.224     -0.000
AIC*n:                    921.266         921.390     -0.124
BIC:                    -4029.663       -4029.539     -0.124
BIC':                     -78.112         -77.988     -0.124

Difference of 0.124 in BIC' provides weak support for current model.

Note: p-value for difference in LR is only valid if models are
nested.
```

Jak widać zarówno na podstawie wartości logarytmu funkcji wiarogodności (bliższy zera), jak i BIC preferowany jest model logitowy.

Test ilorazu wiarogodności jest standardową procedurą od wersji 9.0. Sprawdzimy teraz za jej pomocą czy posiadanie dzieci wpływa na prawdopodobieństwo podjęcia pracy przez kobietę. W tym celu przebadamy:

$$H_0: \quad \beta_{k5} = \beta_{k618} = 0$$

przeciw alteratywie

$$H_1: \quad \beta_{k5} \neq 0 \vee \beta_{k618} \neq 0$$

Aby to zrobić musimy oszacować pełen model

```
. qui probit lfp k5 k618 age wc hc lwg inc
. estimates store probit
```

oraz model z ograniczeniami

```
. qui probit lfp age wc hc lwg inc
. estimates store probit_r
```

a następnie należy porównać oba modele wykorzystując polecenie `lrtest`

```
. lrtest probit probit_r, stats

Likelihood-ratio test                               LR chi2(2)  =     65.97
(Assumption: probit_r nested in probit)             Prob > chi2 =    0.0000

-------------------------------------------------------------------------
     Model |    Obs    ll(null)   ll(model)     df          AIC        BIC
-----------+-------------------------------------------------------------
  probit_r |    753   -514.8732   -485.6786      6    983.3572   1011.102
    probit |    753   -514.8732    -452.695      8    921.3899   958.3824
-------------------------------------------------------------------------
```

Wynik testu jednoznacznie wskazuje, że odrzucana jest hipoteza o tym że model o mniejszej ilości zmiennych zawiera się w modelu o większej ilości zmiennych. Oznacza to, iż ograniczenia nałożone na model są nieprawidłowe.

### 1.2.8 Test prawidłowości formy funkcyjnej

Odpowiednikiem testu RESET dla model szacowanych metodą największej wiarogodności, który również działa dla modelu regresji, jest test sprawdzający poprawność funkcji łączącej obserwację. Jego działanie jest analogiczne do testu RESET w którym jako dodatkowy regresor używany jest kwadrat wartości dopasowanej.

```
. estimates restore probit (results probit are active now)

. linktest

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -451.38098
Iteration 2:   log likelihood = -449.36073
Iteration 3:   log likelihood = -449.34652
Iteration 4:   log likelihood = -449.34652

Probit regression                                 Number of obs   =        753
                                                  LR chi2(2)      =     131.05
                                                  Prob > chi2     =     0.0000
Log likelihood = -449.34652                       Pseudo R2       =     0.1273
------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   .9697123   .0957237    10.13   0.000     .7820974    1.157327
      _hatsq |   .2924195   .1085821     2.69   0.007     .0796026    .5052364
       _cons |  -.0762115   .0586236    -1.30   0.194    -.1911117    .0386887
------------------------------------------------------------------------------
```

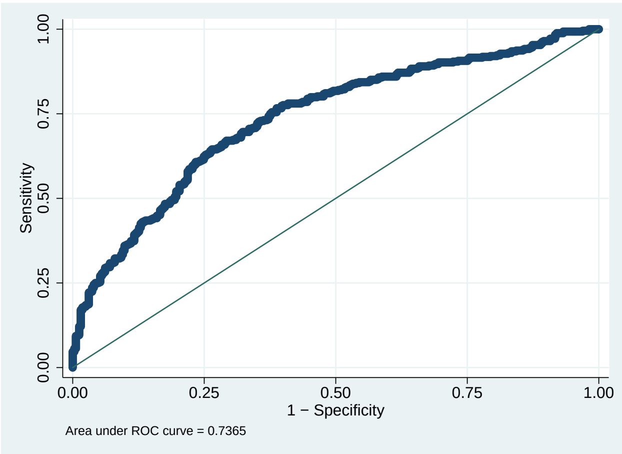
Pierwsze polecenie przywraca wyniki estymacji zachowane pod nazwą probit, polecenie `linktest` sprawdza poprawność formy funkcyjnej. Ponieważ kwadrat wartości dopasowanej jest zmienną istotną w regresji pomocniczej wnioskujemy, że forma funkcyjna modelu nie jest poprawna.

### 1.2.9 Klasyfikacja

Po oszacowaniu modelu wyborów dyskretnych warto jest sprawdzić jak radzi on sobie z klasyfikowaniem kolejnych obserwacji jako *sukcesy* albo *porażki*. W tym celu bada się jego własności dyskryminacyjne, oraz zdolność do prawidłowego przewidywania sukcesów i porażek.

**Krzywa ROC.** Dokładność klasyfikacji danych zależy od stopnia w jakim zbudowany model rozróżnia sukcesy od porażek. Miarą dokładności jest pole zawarte pod krzywą ROC. Pole równe 1 oznacza, że model doskonale dyskryminuje między sukcesem a porażką, wartość równa 0,5 oznacza brak możliwości dyskryminacji.

```
. lroc
```



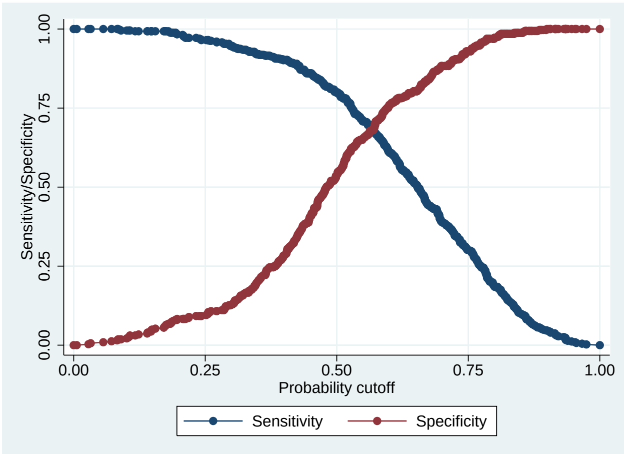

```
Logistic model for lfp

number of observations =      753
area under ROC curve   =   0.7364
```

W analizowany przez nas model charakteryzuje się przeciętną zdolnością do rozróżniania kobiet pracujących i niepracujących.

**Analiza wrażliwości.** Krzywa wrażliwości (ang. *sensivity*) przedstawia prawdopodobieństwo sklasyfikowania obserwacji jako sukces pod warunkiem, że wystąpił on w rzeczywistości $Pr(+ \mid D)$, przy rożnych poziomach podziału wartości dopasowanych na sukcesy i porażki. Krzywa specyficzności (ang.

emphspecifity) reprezentuje prawdopodobieństwo klasyfikacji obserwacji jako porażka, pod warunkiem że ona wystąpi w rzeczywistości $Pr(-|\tilde{}\mathtt{D})$, przy rożnych poziomach podziału wartości dopasowanych na sukcesy i porażki.

```
. lsens
```




**Analiza trafności** Tablica trafności dopasowania porównuje wartości zaobserwowane w rzeczywistości z przewidywaniami modelu. W kolumnach są wartości rzeczywiste. Sukcesy oznaczone są przez `D`, porażki przez `~D`. Wiersze przedstawiają sposób klasyfikacji obserwacji wynikający z dopasowań modelu.

```
. estat clas (lstat w Stata8)

Probit model for lfp

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       342           150  |        492
     -     |        86           175  |        261
-----------+--------------------------+-----------
   Total   |       428           325  |        753

Classified + if predicted Pr(D) >= .5 True D defined as lfp != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   79.91%
Specificity                     Pr( -|~D)   53.85%
Positive predictive value       Pr( D| +)   69.51%
Negative predictive value       Pr(~D| -)   67.05%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   46.15%
```

```
False - rate for true D         Pr( -| D)   20.09%
False + rate for classified +   Pr(~D| +)   30.49%
False - rate for classified -   Pr( D| -)   32.95%
--------------------------------------------------
Correctly classified                        68.66%
--------------------------------------------------
```

Trafność dopasowania zależy zależy od przyjętego poziomu, który rozgranicza wartości dopasowane w modelu na sukcesy i porażki. Standardowo przyjęta wartością jest (0,5). Jeżeli zostanie ustalony poziom wyznaczony przez punkt przecięcia krzywej wrażliwości i specyficzności to zostanie zmaksymalizowana liczba poprawnie sklasyfikowanych obserwacji.

```
. lstat, cutoff(0.52)

Probit model for lfp

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       330           129  |        459
     -     |        98           196  |        294
-----------+--------------------------+-----------
   Total   |       428           325  |        753

Classified + if predicted Pr(D) >= .52 True D defined as lfp != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   77.10%
Specificity                     Pr( -|~D)   60.31%
Positive predictive value       Pr( D| +)   71.90%
Negative predictive value       Pr(~D| -)   66.67%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   39.69%
False - rate for true D         Pr( -| D)   22.90%
False + rate for classified +   Pr(~D| +)   28.10%
False - rate for classified -   Pr( D| -)   33.33%
--------------------------------------------------
Correctly classified                        69.85%
--------------------------------------------------
```

Wybór miar na które należy zwrócić szczególną uwagę zależy od konkretnego celu badania. Jeżeli model jest używany do kontroli jakości leków to chcemy minimalizować przypadki w których pacjent poddany kuracji umrze, czyli będziemy minimalizować „`False - rate for true D`”. Jeżeli chcemy uniknąć fałszywej klasyfikacji, np przy identyfikacji przestępców szczególną uwagę powinniśmy zwrócić na minimalizowanie „`False + rate for true D`”.

## 1.3 Modele wielowartościowe

W badaniach ekonomicznych i społecznych często odpowiedzi na pytania są kodowane za pomocą skali Likerta zgodności z zaproponowanym stwierdzeniem. Wartości tej skali są przyjmowane arbitralne, zazwyczaj od 1 do 5, lub od 1 do 7. Z uwagi na arbitralne ustalenie wartości jakie przyjmuje zmienna nie posiada ona interpretacji ilościowej, ponadto z reguły wartości pośrednie nie posiadają interpretacji. Z tego powodu nie powinno się traktować takich zmiennych identycznie jak traktowane są zmienne kardynalne.

Gdy chcemy zmienną o wielu wartościach użyć jako zmienną zależną modelu to wygodnym narzędziem ekonometrycznym są uogólnione modele wyborów dyskretnych.

### 1.3.1 Modele wielomianowe

Załóżmy, że jednostka dokonuje wyboru spośród $K$ dostępnych alternatyw ($K > 2$), dla $K = 2$ problem redukuje się do modelu z binarną zmienna zależną. Jeżeli założymy, że jednostki są racjonalne, to osoba $i$ wybierze alternatywę $j$, gdy powiązana z nią użyteczność będzie wyższa niż użyteczność uzyskana z wyboru każdej z pozostałych alternatyw.

$$Pr(y_j = j) = Pr(U_j > U_k \quad \forall j \neq k)$$

Inne podejście do modelowania wyborów dyskretnych zakłada, że za wybór odpowiada nieobserwowana zmienna ciągła $y^*$, o której zakłada się, że ma znany rozkład, zazwyczaj normalny lub logistyczny. Obserwowane jest zmienna dyskretna $y$ przyjmująca wartości $1 \dots K$. Wartości tej zmiennej są związane z wartościami zmiennej $y^*$.

**Niezależność niezwiązanych alternatyw (IIA)** Ważnym założeniem modeli wielomianowych jest niezależność wykluczających się alternatyw. W przypadku pojawienia się nowej alternatywy, wszystkie prawdopodobieństwa dotychczas uwzględnionych alternatyw powinny się tak dostosować, aby ilorazy szans wystąpienia danej wartości pomiędzy wszystkimi możliwymi parami kategorii zmiennej objaśnianej pozostały bez zmian. By lepiej zobrazować to założenie przedstawię przykład, który opisał McFadden.

Przypuśćmy, że w pewnym mieście ludzie dojeżdżający do pracy mają do wyboru jazdę samochodem, albo autobusem. Dwie trzecie z nich wybiera samochód. Przypuśćmy, że powstaje nowy "rodzaj" autobusu. Intuicyjnie rzecz ujmując dwie trzecie populacji dalej powinno wybierać samochód jako środek transportu, a pozostali powinni podzielić się pomiędzy dwa rodzaje autobusu. Ale by zachować założenie o niezależności wykluczających się alternatyw, tylko połowa ludzi powinna używać własnych samochodów od momentu wprowadzenia drugiego rodzaju autobusu. Ponieważ, do momentu uruchomienia nowego autobusu było:

$$\frac{Pr(samochod)}{Pr(autobus)} = \frac{2}{1}$$

to po wprowadzeniu drugiego rodzaju autobusu, ta proporcja powinna być zachowana. Ale jej zachowanie powoduje że prawdopodobieństwo wyboru samochodu jako środka transportu spadnie z 2/3 do 1/2, bowiem

$$\frac{Pr(samochod)}{Pr(autobus_1)} = \frac{2}{1} \quad \frac{Pr(samochod)}{Pr(autobus_2)} = \frac{2}{1}$$

wobec tego $P(samochd) = \frac{1}{2}$, $P(autobus_1) = P(autobus_2) = \frac{1}{4}$.

Modele wielomianowe zakładają niezależność wykluczających się alternatyw. Z tego powodu powinny być używane jedynie w przypadku, gdy kategorie zmiennej objaśnianej są niezależne, oraz rozróżnialne dla jednostki podejmującej decyzję lub dokonującej wyboru. Dodatkowym warunkiem poprawności modelu jest, że różnica pomiędzy kategoriami zmiennej objaśnianej powinna mieć istotny wpływ na decyzję o wyborze. Założenie o niezależności składników losowych powoduje, że modele wielomianowe są poprawne tylko w przypadku braku podobieństw pomiędzy dostępnymi alternatywami.

### 1.3.2 Przykład 2.

Dane do przykładu pochodzą z amerynańskiego rynku obligacji komercyjnych. Zbiór liczy 98 obserwacji i zawiera informacje firmach. Zmienną zależną jest rating obligacji `rating83` od AAA do C, który zakodowany jest jako wartość całkowita, im wyższa tym wyższy rating. Jego poziom jest tłumaczony za pomocą wskaźnika dochód do wartości `ia83` (ang. *income-to-asset ratio*), oraz zmianą poziomu tego wskaźnika między rokiem 1982 a 1983, zmienna `dia`.

Dane są dostępne w internecie. Wystarczy w Stacie wpisać

```
. use http://www.stata-press.com/data/imeus/panel84extract, clear
```

Zmienna oryginalna `rating` dla pewnych kategorii zawiera małą liczbę obserwacji.

```
. tab rating83

   rating83 |      Freq.     Percent        Cum.
------------+-----------------------------------
        AAA |         29       29.59       29.59
```

```
         AA |          2        2.04       31.63
          A |         13       13.27       44.90
        BAA |         28       28.57       73.47
         BA |         16       16.33       89.80
          B |          7        7.14       96.94
          C |          3        3.06      100.00
------------+-----------------------------------
      Total |         98      100.00
```

Gdybyśmy chcieli użyć zmiennej `rating83` jako objaśnianej to dla wyboru między ratingiem C a ratingiem B dysponowalibyśmy tylko 10 obserwacjami. W modelu są 2 zmienne, stała, należy oszacować wariancję, czyli zostaje tylko 6 stopnii swobody. Powodowało by to niską precyzję oszacowań.

Z tego powodu przed przystąpieniem do szacowania parametrów modelu należy ją przekształcić w taki sposób, aby w każda kategoria przekształconej zmiennej zawierała podobną ilość obserwacji. Taka operacja ułatwia, a czasami wręcz umożliwia oszacowanie parametrów modelu.

```
. tab rating83c

       Bond |
    rating, |
       1983 |      Freq.     Percent        Cum.
------------+-----------------------------------
     BA_B_C |         26       26.53       26.53
        BAA |         28       28.57       55.10
       AA_A |         15       15.31       70.41
        AAA |         29       29.59      100.00
------------+-----------------------------------
      Total |         98      100.00
```

Na początku potraktujmy rating jako zmienną bez ustalonej hierarchii[2]. Wobec tego aby zbudować model użyjemy modelu wielomianowego.

```
. mlogit rating83c ia83 dia

Iteration 0:   log likelihood = -133.04224
Iteration 1:   log likelihood = -119.74382
Iteration 2:   log likelihood = -118.09549
Iteration 3:   log likelihood = -118.00312
Iteration 4:   log likelihood = -118.00239

Multinomial logistic regression                   Number of obs   =         98
                                                  LR chi2(6)      =      30.08
                                                  Prob > chi2     =     0.0000
```

[2]Przyjmujemy takie założenie wyłącznie w celach szkoleniowych.

```
Log likelihood = -118.00239                       Pseudo R2       =     0.1130
------------------------------------------------------------------------------
   rating83c |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
BA_B_C       |
        ia83 | -.1303426    .0489834    -2.66   0.008    -.2263484   -.0343369
         dia |  .1542639    .0735878     2.10   0.036     .0100345    .2984934
       _cons |  .9820093    .4966503     1.98   0.048     .0085926    1.955426
-------------+----------------------------------------------------------------
BAA          |
        ia83 | -.0454026    .0441004    -1.03   0.303    -.1318378    .0410326
         dia |  .1240656    .0709813     1.75   0.080    -.0150552    .2631865
       _cons |  .4024031    .5154444     0.78   0.435    -.6078494    1.412656
-------------+----------------------------------------------------------------
AA_A         |
        ia83 |  .1464149    .0593538     2.47   0.014     .0300836    .2627462
         dia |  .1748879    .0910389     1.92   0.055    -.0035451    .3533209
       _cons | -2.909247    .9744167    -2.99   0.003    -4.819069   -.9994255
------------------------------------------------------------------------------
(rating83c==AAA is the base outcome)
```

Wyniki dla każdej kategorii są liczone w odniesieniu do poziomu bazowego. Stata jako poziom odniesienia ustaliła rating AAA, ponieważ dla tej kategorii dostępną jest największa liczba obserwacji. Za pomocą opcji `baseoutcome(#)`, gdzie # oznacza numer alternatywy, można kontrolować poziom odniesienia. W wierszu pierwszym są współczynniki porównujące rating BA_B_C z AAA, w drugim porównany jest rating BAA z AAA, a w trzecim AA_A. Zbliżone wyniki można uzyskać szacując osobno modele logitowe dla każdej kategorii. Wyniki nie będą takie same, bowiem w przypadku logitu funkcja wiarogodności zależy od 4 parametrów, a w przypadku wielomianowego logitu od 10.

Analogiczne wyniki można uzyskać szacując wielomianowy model probitowy

```
. mprobit rating83c ia83 dia

Iteration 0:   log likelihood = -117.80861
Iteration 1:   log likelihood = -117.62741
Iteration 2:   log likelihood =  -117.6271
Iteration 3:   log likelihood =  -117.6271

Multinomial probit regression                     Number of obs   =         98
                                                  Wald chi2(6)    =      21.56
Log likelihood =  -117.6271                       Prob > chi2     =     0.0015

------------------------------------------------------------------------------
   rating83c |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
BA_B_C       |
```

```
        ia83 | -.0911987   .0331999    -2.75  0.006    -.1562694   -.026128
         dia |  .0995189    .050054     1.99  0.047     .0014149    .197623
       _cons |  .6635147   .3480602     1.91  0.057    -.0186707     1.3457
-------------+----------------------------------------------------------------
BAA          |
        ia83 | -.0284042   .0307741    -0.92  0.356    -.0887203   .0319118
         dia |  .0834583   .0505054     1.65  0.098    -.0155304   .1824471
       _cons |  .2136007   .3606114     0.59  0.554    -.4931847   .9203861
-------------+----------------------------------------------------------------
AA_A         |
        ia83 |  .1113513   .0418002     2.66  0.008     .0294244   .1932782
         dia |  .1167499   .0636459     1.83  0.067    -.0079937   .2414935
       _cons | -2.143089   .6606959    -3.24  0.001    -3.438029   .8481486
------------------------------------------------------------------------------
```

Porównując wartość logarytmu funkcji wiarogodności i wartość statystyki LR łatwo zauważyć, że wielomianowy model logitowy jest lepiej dopasowany do danych empirycznych.

Tak jak w modelu logitowym wartości współczynników nie mają interpretacji ekonomicznej. Aby im ją nadać należy przedstawić wyniki w postaci ilorazów szans (ryzyk) (ang. *relative risk ratio*)

```
mlogit, rrr

Multinomial logistic regression                   Number of obs   =         98
                                                  LR chi2(6)      =      30.08
                                                  Prob > chi2     =     0.0000
Log likelihood = -118.00239                       Pseudo R2       =     0.1130

------------------------------------------------------------------------------
   rating83c |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
BA_B_C       |
        ia83 |  .8777946   .0429974    -2.66  0.008     .7974402    .9662459
         dia |  1.166799   .0858622     2.10  0.036     1.010085    1.347827
-------------+----------------------------------------------------------------
BAA          |
        ia83 |  .9556127   .0421429    -1.03  0.303     .8764831    1.041886
         dia |   1.13209   .0803573     1.75  0.080     .9850576    1.301069
-------------+----------------------------------------------------------------
AA_A         |
        ia83 |  1.157676   .0687125     2.47  0.014     1.030541    1.300497
         dia |  1.191113   .1084376     1.92  0.055     .9964612    1.423788
------------------------------------------------------------------------------
(rating83c==AAA is the base outcome)
```

Im wyższa była wartość wskaźnika dochód do wartości, tym większa szansa na wyższy rating, natomiast wpływ zmiany tego wskaźnika jest ujemny.

**Testowanie istotności** Na wydruku ze Staty widzimy, że zmienne są łącznie istotne (statystyka LR). Natomiast warto jest przetestować istotność poszczególnych zmiennych. Można zrobić to dwoma metodami. Testem Walda

```
. test ia83

 ( 1)  [BA_B_C]ia83 = 0
 ( 2)  [BAA]ia83 = 0
 ( 3)  [AA_A]ia83 = 0

           chi2(  3) =   17.07
         Prob > chi2 =    0.0007

. test dia

 ( 1)  [BA_B_C]dia = 0
 ( 2)  [BAA]dia = 0
 ( 3)  [AA_A]dia = 0

           chi2(  3) =    5.82
         Prob > chi2 =    0.1208
```

Na podstawie tego testu odrzucamy hipotezę o łącznej nieistotności zmiennej `ia83`, natomiast zmienna `dia` jest statystycznie nieistotna.

Alternatywnym testem jest test ilorazu wiarogodności. W uzyskaniu jego wyników, jak również innych testów, bardzo przydatny jest dodatkowy pakiet `mlogtest`.

```
. mlogtest, l w

**** Likelihood-ratio tests for independent variables

 Ho: All coefficients associated with given variable(s) are 0.

    rati~83c |       chi2   df   P>chi2
-------------+-------------------------
        ia83 |     23.935    3    0.000
         dia |      6.741    3    0.081
---------------------------------------

**** Wald tests for independent variables

 Ho: All coefficients associated with given variable(s) are 0.

    rati~83c |       chi2   df   P>chi2
-------------+-------------------------
        ia83 |     17.068    3    0.001
```

```
        dia |      5.819    3    0.121
---------------------------------------
```

Wartości statystyk nieznacznie różnią się w obu testach, jednak dają one takie same konkluzję.

**Testowanie niezależności niezwiązanych alternatyw** W pakiecie Stata za zaimplementowane dwa testy sprawdzające założenie o niezależności niezwiązanych alternatyw. W obu hipotezą zerową jest niezależność niezwiązanych alternatyw. Oba mają podobną konstrukcję i porównują oszacowania przy pełnym zestawie alternatyw i pominięciu jednej z nich.

Test Hausmana opiera się o statystykę Walda.

```
. mlogtest, h sm
**** Hausman tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |      chi2   df   P>chi2   evidence
---------+------------------------------------
  BA_B_C |    -1.834    6    1.000   for Ho
     BAA |    -5.012    6    1.000   for Ho
    AA_A |    -1.956    6    1.000   for Ho
----------------------------------------------
```

Test Small'a i Hsiao bazuje na statystyce ilorazu wiarogodności.

```
 **** Small-Hsiao tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |  lnL(full)  lnL(omit)    chi2   df   P>chi2   evidence
---------+---------------------------------------------------------
  BA_B_C |    -35.112    -33.256   3.713    3    0.294   for Ho
     BAA |    -29.480    -26.636   5.688    3    0.128   for Ho
    AA_A |    -36.001    -34.593   2.816    3    0.421   for Ho
-------------------------------------------------------------------
```

Oba testy wskazują, że założenie o niezależności alternatyw jest spełnione. Jest to rzadko spotykany przypadek. Szczególnie w dużych próbach testy mogą dawać przeczące sobie wyniki.

Warto również sprawdzić czy pewnych kategorii zmiennej zależnej nie da się połączyć w jedną.

```
. mlogtest, c lrc
 **** Wald tests for combining outcome categories
```

```
 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).

Categories tested |      chi2   df   P>chi2
------------------+------------------------
  BA_B_C-     BAA |     3.287    2    0.193
  BA_B_C-    AA_A |    17.198    2    0.000
  BA_B_C-     AAA |     8.067    2    0.018
     BAA-    AA_A |    10.214    2    0.006
     BAA-     AAA |     3.212    2    0.201
    AA_A-     AAA |    10.273    2    0.006
-------------------------------------------

**** LR tests for combining outcome categories

 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).

Categories tested |      chi2   df   P>chi2
------------------+------------------------
  BA_B_C-     BAA |     3.534    2    0.171
  BA_B_C-    AA_A |    24.973    2    0.000
  BA_B_C-     AAA |     9.476    2    0.009
     BAA-    AA_A |    13.573    2    0.001
     BAA-     AAA |     3.392    2    0.183
    AA_A-     AAA |    14.280    2    0.001
-------------------------------------------
```

Oba testy dają podobne wyniki wskazując, że kategorię BAA można połączyć z kategorią BA_B_C, oraz można ją połączyć z kategorią AAA.

Wszystkie testy można wywołać poleceniem

```
. mlogtest, all
```

Na koniec zapamiętajmy wyniki oszacowań

```
. qui mlogit rating83c ia83 dia
. est store mlogit
```

### 1.3.3 Modele uporządkowane

Czasami zmienna zależna o charakterze nominalnym posiada naturalną hierarchię. Wobec tego modelując zjawisko można i wskazane jest taką informację wykorzystać.

Przystępując do analizy danych o ratingu w poprzednim punkcie pominęliśmy fakt uszeregowania zmiennej zależnej.

```
. ologit rating83c ia83 dia

Iteration 0:   log likelihood = -133.04224
Iteration 1:   log likelihood = -127.30126
Iteration 2:   log likelihood = -127.27148
Iteration 3:   log likelihood = -127.27146

Ordered logistic regression                       Number of obs   =         98
                                                  LR chi2(2)      =      11.54
                                                  Prob > chi2     =     0.0031
Log likelihood = -127.27146                       Pseudo R2       =     0.0434

------------------------------------------------------------------------------
   rating83c |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ia83 |   .0939166   .0296196     3.17   0.002     .0358633    .1519699
         dia |  -.0866925   .0449789    -1.93   0.054    -.1748496    .0014646
-------------+----------------------------------------------------------------
       /cut1 |  -.1853053   .3571432                     -.8852931    .5146825
       /cut2 |   1.185726   .3882098                      .4248489    1.946603
       /cut3 |   1.908412   .4164895                      1.092108    2.724717
------------------------------------------------------------------------------
```

Podobne wyniki uzyskamy szacując uszeregowany model probitowy

```
. oprobit rating83c ia83 dia

Iteration 0:   log likelihood = -133.04224
Iteration 1:   log likelihood = -127.87966
Iteration 2:   log likelihood = -127.87756

Ordered probit regression                         Number of obs   =         98
                                                  LR chi2(2)      =      10.33
                                                  Prob > chi2     =     0.0057
Log likelihood = -127.87756                       Pseudo R2       =     0.0388

------------------------------------------------------------------------------
   rating83c |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ia83 |   .0512509   .0167257     3.06   0.002     .0184691    .0840327
         dia |  -.0496082   .0254406    -1.95   0.051    -.0994709    .0002545
-------------+----------------------------------------------------------------
       /cut1 |  -.1877442   .2048421                     -.5892274     .213739
       /cut2 |   .6344613   .2156936                      .2117095    1.057213
       /cut3 |   1.064523   .2231879                      .6270823    1.501963
------------------------------------------------------------------------------
```

Tak jak w przypadku modelu wielomianowego model logitowy wydaje się być lepiej dopasowany do danych ze względu na wyższą wartość logarytmu funkcji wiarogodności oraz statystyki LR.

Uporządkowany model logitowy może być traktowany w pewnym przybliżeniu jako model wielomianowy z narzuconymi ograniczeniami. Wobec tego można porównać oba modele.

```
. lrtest ologit mlogit, force

Likelihood-ratio test                                 LR chi2(4)=     18.54
(Assumption: ologit nested in mlogit)                 Prob > chi2 =    0.0010
```

Test wskazuje, że model uporządkowany nie jest zagnieżdżony w modelu wielomianowym, wobec tego wnosi dodatkową informację.

## 1.4 Modele nielosowej próby

W badaniach ekonometrycznych próba, którą dysponujemy nie zawsze powstaje w sposób czysto losowy. Przeanalizujemy dwa przypadki odejścia od założenia o jej losowym charakterze. Pierwszy z nich będzie analizował obcięcie, bądź ocenzurowanie rozkładu zmiennej zależnej, drugi nie w pełni losowy mechanizm doboru obserwacji do próby i jego konsekwencje.

### 1.4.1 Tobit

Model próby uciętej powstaje w efekcie obcięcia rozkładu z którego próba jest losowana, lub ocenzurowania wartości zmiennej zależnej. W przypadku ucięcia niedostępne są informacje dotyczące zarówno zmiennej zależnej, jak również zmiennych objaśniających. Ocenzurowanie jest innym, często spotykanym w praktyce badawczej mechanizmem. Ogranicza on zakres zmienności zmiennej zależnej. Ocenzurowanie występuje, gdy zmiennej zależnej jest arbitralnie przypisywana pewna wartość, gdy prawdziwa wartość przekroczy pewien ustalony poziom.

Rozkład obcięty jest to rozkład zmiennej losowej w przypadku, gdy przekracza ona pewną wartość. Wpływ ucięcia na momenty zmiennej losowej wydaje się oczywisty. Wartość oczekiwana (średnia) oddala się od punktu odcięcia, a wariancja zmiennej zmniejsza się.

Rozkład zmiennej uciętej w punkcie $a$ dany jest przez:

$$f(x \mid x > a) = \frac{f(x)}{Pr(X > a)}$$

gdzie $X$ jest ciągła zmienną losową, a $f(x)$ jej funkcją gęstości. W zastosowaniach najczęściej używa się obciętego rozkładu normalnego. W takim przypadku:

$$Pr(X > a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right) = 1 - \Phi(\alpha)$$

gdzie $\mu, \sigma$ to średnia i odchylenie standardowe rozkładu normalnego, $\Phi(.)$ jego dystrybuanta,a $\alpha = \frac{a-\mu}{\sigma}$.

**Lemat 3** *Jeśli zmienna losowa $X \sim N(\mu, \sigma^2)$ oraz a jest stałą, wtedy:*

$$E(X \mid \textit{obcięcie}) = \mu + \sigma\lambda(\alpha)$$

$$Var(X \mid \textit{obcięcie}) = \sigma^2[1 - \delta(\alpha)]$$

*gdzie:*

$$\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)}$$

$$\lambda(\alpha) = \frac{-\phi(\alpha)}{\Phi(\alpha)} \qquad x > a$$

$$\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha] \qquad x < a$$

Funkcja $\lambda(\alpha)$ nazywana jest odwróconym ilorazem Millsa (*inverse Mills ratio*)

Problem ocenzurownia polega na tym, że pewna część wartości zmiennej zależnej od pewnego jej poziomu jest niedostępna. W takim przypadku, jeżeli zmienna zależna przekracza pewną wartość, to jako wynik zawsze dostajemy tą samą liczbę. Zmienna może być ocenzurowana z dołu lub z góry, bądź z obu stron jednocześnie. Ekonomicznymi przykładami ocenzurowania zmiennej zależnej są: zakup przez gospodarstwa domowe dóbr trwałych, liczba przepracowanych godzin przez kobiety, czy wydatki na urlopy wypoczynkowe. W każdym z wymienionych przykładów wiele obserwacji przyjmuje wartość zero.

Dane mogą być traktowane jako realizacja nieobserwowanego procesu $y_i^*$, dla którego jedynie obserwujemy:

$$y_i = \begin{cases} y_i^* & \text{dla} \quad y_i \geq 0 \\ 0 & \text{dla} \quad y_i < 0 \end{cases} \tag{10}$$

Stosowanie modeli zmiennej ocenzurowanej jest zalecane, gdy znacząca część obserwacji, powiedzmy więcej niż 5 %, leży na granicy przedziału w którym obserwujemy wartość zmiennej.

Forma modelu ekonometrycznego dla zmiennej ocenzurowanej została zaproponowana przez Tobina w 1958 roku. Zaproponował on model

$$y_i^* = X_i'\beta + \varepsilon \tag{11}$$

$$\varepsilon \sim_{IID} N(0, \sigma^2)$$

w którym obserwowana zmienna $y_i$ jest generowana przez równanie (10).

Proste oszacowanie (11) przez MNK prowadzi do uzyskania obciążonych estymatorów dla populacji bez wprowadzenia poprawki, ze względu na nielosowość cenzurowania. Dzieje się tak, ponieważ $E(Y_i \mid X) = X'\beta$, zamiast $E(Y_i^* \mid X) = X'\beta$. Oszacowanie równania regresji na dostępnej podpróbie nie rozwiązuje problemu, bowiem uzyskane oszacowania współczynników będą obciążone w kierunku zera lub niedokładne. Również wariancja będzie niedoszacowana.

Składnik losowy równania modelu ma rozkład $\varepsilon \sim N(0, \sigma^2)$. Wobec tego zmienna objaśniana ma rozkład:

$$(Y_i \mid X_i) \sim N(X_i'\beta, \sigma^2)$$

dla wartości zmiennej objaśnianej większych od punktu obcięcia, $Y_i$ ma rozkład:

$$E[Y_i \mid Y_i > a] = X_i'\beta + \sigma \frac{\phi(\frac{a - X_i'\beta}{\sigma})}{1 - \phi(\frac{a - X_i'\beta}{\sigma})}$$

Jeżeli punkt obcięcia $a$ znormalizujemy do zera, to rozkład upraszcza się do:

$$Pr(y_i > y \mid X_i) = \Phi\left(\frac{X_i\beta}{\sigma}\right) = \int_0^y \frac{1}{\sigma}\phi\left(\frac{z - X_i\beta}{\sigma}\right) dz \tag{12}$$

Zgodne estymatory otrzymywane są metodą największej wiarogodności. Prawdopodobieństwo ocenzurowania wynosi:

$$Pr(y_i = 0 \mid X_i) = Pr(y_i^* < 0 \mid X_i) = Pr(X_i'\beta + \varepsilon_i < 0 \mid X_i) \tag{13}$$

$$Pr(y_i = 0 \mid X_i) = Pr\left(\frac{\varepsilon_i}{\sigma} < -\frac{X_i'\beta}{\sigma} \mid X_i\right) = \Phi\left(-\frac{X_i'\beta}{\sigma}\right)$$

Jeżeli zestawimy (12) oraz (13) to możemy zapisać funkcję wiarogodności dla modelu tobitowego.

$$L(\beta) = \Phi\Big(-\frac{X_i'\beta}{\sigma}\Big)^{\sum \mathbb{I}(y=0)} \Big[\frac{1}{\sigma}\phi\Big(\frac{z - X_i\beta}{\sigma}\Big)\Big]^{\sum \mathbb{I}(y>0)} \tag{14}$$

Gdzie $\mathbb{I}$ oznacza funkcję indykatorową. Estymator wektora $\beta$ otrzymujemy przez maksymalizację logarytmu funkcji wiarogodności.

### 1.4.2 Przykład 3.

Rozważmy próbę pracujących kobiet. Z naturalnych powodów liczba godzin pracy jest z dołu ograniczona przez 0.

```
. use http://www.stata-press.com/data/imeus/laborsub, clear

. des

Contains data from
http://www.stata-press.com/data/imeus/laborsub.dta
  obs:           250
 vars:             6                          25 Sep 2004 18:36
 size:         2,750 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
lfp             byte   %9.0g                  1 if woman worked in 1975
whrs            int    %9.0g                  Wife's hours of work
kl6             byte   %9.0g                  # of children younger than 6
k618            byte   %9.0g                  # of children between 6 and 18
wa              byte   %9.0g                  Wife's age
we              byte   %9.0g                  Wife's educational attainment
-------------------------------------------------------------------------------
```

Co się stanie jeżeli pominiemy problem ucięcia zmiennej i oszacujemy model regresji na nieuciętych obserwacjach .

```
. reg  whrs kl6 k618 wa we if whrs>0

   Source |       SS       df       MS              Number of obs =     150
----------+------------------------------           F(  4,   145) =    2.80
    Model |  7326995.15     4  1831748.79           Prob > F      =  0.0281
 Residual |  94793104.2   145  653745.546           R-squared     =  0.0717
----------+------------------------------           Adj R-squared =  0.0461
    Total |   102120099   149  685369.794           Root MSE      =  808.55

------------------------------------------------------------------------------
     whrs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
      kl6 | -421.4822   167.9734    -2.51   0.013    -753.4748   -89.48953
     k618 | -104.4571   54.18616    -1.93   0.056    -211.5538    2.639668
       wa | -4.784917   9.690502    -0.49   0.622     -23.9378    14.36797
       we |  9.353195   31.23793     0.30   0.765    -52.38731     71.0937
    _cons |  1629.817   615.1301     2.65   0.009     414.0371    2845.597
------------------------------------------------------------------------------
```

Jeżeli możemy przyjąć założenie, że składnik losowy w populacji ma rozkład normalny, to możemy uzyskać poprawne oszacowania dla uciętej próby wykorzystując polecenie `truncreg` z odpowiednią opcją `ll`(·) dla ucięcia z dołu , oraz `ul`(·) dla ucięcia z góry. Przyjmując założenie o normalności rozkładu na podstawie oszacowań możemy wnioskować o całej populacji.

```
. truncreg whrs kl6 k618 wa we, ll(0)
(note: 100 obs. truncated)

Fitting full model:

Iteration 0:   log likelihood = -1205.6992
Iteration 1:   log likelihood = -1200.9873
Iteration 2:   log likelihood = -1200.9159
Iteration 3:   log likelihood = -1200.9157
Iteration 4:   log likelihood = -1200.9157

Truncated regression
Limit:   lower =          0                     Number of obs =    150
         upper =       +inf                     Wald chi2(4)  =  10.05
Log likelihood = -1200.9157                     Prob > chi2   = 0.0395

------------------------------------------------------------------------------
     whrs |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
eq1       |
      kl6 | -803.0042   321.3614    -2.50   0.012    -1432.861   -173.1474
     k618 |  -172.875   88.72898    -1.95   0.051    -346.7806    1.030578
       wa | -8.821122   14.36848    -0.61   0.539    -36.98283    19.34059
       we |  16.52873   46.50375     0.36   0.722    -74.61695    107.6744
    _cons |   1586.26    912.355     1.74   0.082    -201.9233    3374.442
----------+-------------------------------------------------------------------
sigma     |
    _cons |  983.7262   94.44303    10.42   0.000     798.6213    1168.831
------------------------------------------------------------------------------
```

Jak widać z porównania wyników obu oszacowań, w modelu regresji zgodnie z przewidywaniami oszacowane wartości parametrów i wariancji są znacznie niższe. Estymatory parametrów są prawie dwukrotnie niższe, a odchylenia standardowego `sigma`=984, podczas gdy RMSE=809. Natomiast wpływ na istotność zmiennych jest niewielki.

Jeżeli przyjmiemy, że każda kobieta, która chce pracować robi to, to problem z którym mamy do czynienia to nie ucięcie rozkładu a ocenzurowanie. Dla tych kobiet czas pracy nie jest dostępny, ale w skutek arbitralnej decyzji badacza został ustalony na zero.Prowadzi to do modelu tobitowego. W tym modelu zakłada się, że te same zmienne determinują fakt nieocenzurowania i wartość nieocenzurowanej obserwacji.

```
. tobit whrs kl6 k618 wa we, ll(0)

Tobit regression                                  Number of obs   =        250
                                                  LR chi2(4)      =      23.03
                                                  Prob > chi2     =     0.0001
```

```
Log likelihood = -1367.0903                        Pseudo R2       =     0.0084

------------------------------------------------------------------------------
        whrs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         kl6 | -827.7657   214.7407    -3.85   0.000    -1250.731   -404.8008
        k618 | -140.0192   74.22303    -1.89   0.060    -286.2129    6.174547
          wa | -24.97919   13.25639    -1.88   0.061    -51.08969    1.131317
          we |  103.6896   41.82393     2.48   0.014     21.31093    186.0683
       _cons |  589.0001   841.5467     0.70   0.485    -1068.556    2246.556
-------------+----------------------------------------------------------------
      /sigma |  1309.909   82.73335                      1146.953    1472.865
------------------------------------------------------------------------------
  Obs. summary:        100  left-censored observations at whrs<=0
                       150     uncensored observations
                         0 right-censored observations
```

Model Tobitowy jest połączeniem modelu liniowego i probitowego. Wobec tego poważny problem dla uzyskania poprawnych oszacowań stanowi heteroscedastyczność. Dlatego częściej stosuje się model Heckamana, który jest bardziej elastyczny i dodatkowo pozwala na to by różne czynniki determinowały ocenzurowanie i poziom zjawiska.

### 1.4.3 Heckit

Zjawisko samoselekcji, nielosowej selekcji lub obcięcia próby zostało szeroko opisane we współczesnej literaturze ekonomicznej, zarówno teoretycznej jak i empirycznej. Znaczna część prac w tej dziedzinie dotyczy ekonomii rynku pracy. Pomimo tego, takie same techniki mogą być z powodzeniem stosowane na innych polach, np. długie szeregi zwrotów z papierów wartościowych są samoselekcjonującą się próbą ze względu na obciążenie związane z przetrwaniem (*survivorship bias*), w medycynie przy badaniach nad nowymi lekami występuje obciążenie związane z wycieraniem się próby (*attrition bias*).

Próba którą dysponuje badacz jest reprezentatywna dla całej populacji ale obserwacje dla zmiennej zależnej są ucięte zgodnie z regułą, które błędy są skorelowane z błędami równania analizowanego zjawiska.

**Przypadkowe obcięcie** W przypadku występowania przypadkowego ucięcia próba jest podzbiorem populacji. Zmienna zależna jest obserwowana, wyłącznie wtedy gdy wartość innej zmiennej wskaże, że obserwacja zostanie wyselekcjonowana.

### 1.4.4 Przykład 4.

W badaniu „bardziej zamożnych amerykanów” okazało się, że ich przeciętny dochód to 142.000 $ rocznie. Ludzie podlegający badaniu mieli albo conajmniej 100.000 $ rocznego dochodu, albo dysponowali majątkiem powyżej 500.000 $ (nie wliczając wartości domu). Przypuśćmy, że badanie dochodów bazowało tylko na jednostkach. których dochód przekraczał 500.000 $. Taka selekcja jest formą obcięcia próby, ale sposób selekcji nie koniecznie obcina wszystkie obserwacje o niskich dochodach. Jednak nadal możemy przypuszczać, że przeciętnie jednostki posiadające duży majątek równocześnie będą miały wysokie dochody. Z tego powodu przeciętny dochód w tej subpopulacji nie będzie dobrym przybliżeniem dochodu przeciętnego amerykanina. Dane w takim badaniu są wybrane nielosowo lub są przypadkowo obcięte.

Studia nad nielosowością próby analizują szkodliwe efekty doboru próby na właściwości tradycyjnych estymatorów, np. otrzymanych metodą najmniejszych kwadratów. Dzięki nim powstały alternatywne metody estymacji i wiele modeli empirycznych. W wielu przypadkach analiza doprowadziła do nowej interpretacji wcześniej uzyskanych wyników.

### 1.4.5 Model selekcji

Proces budowy podstawowej wersji modelu nielosowego doboru rozpoczyna się od równania badanego zjawiska, które zazwyczaj jest tradycyjnym równaniem regresji

$$y_i^* = X_i'\beta + \varepsilon_i \tag{15}$$

Obserwujemy wartości zmiennej wynikowej $y_i$ tylko dla wyselekcjonowanej części próby. Drugim elementem jest równanie opisujące mechanizm wyboru obserwacji do próby. Dobór zależy od obserwowanych charakterystyk jednostek

$$d_i^* = Z_i'\gamma + \nu_i \tag{16}$$

Zmienna zależna $d_i^*$ jest wskaźnikiem podjętej decyzji o zaliczeniu obserwacji do próby.

W celu uproszczenia postaci funkcyjnej i interpretacji parametrów równania (16) definiuje się wartość progową zmiennej decyzyjnej $\bar{d}_i$. Wykorzystuje się ją do zastąpienia nieobserwowanej wartości zmiennej decyzyjnej indykatorową zmienną zero-jedynkową

$$d_i = \begin{cases} 1 \text{ gdy } d_i^* \geq \bar{d}_i \\ 0 \text{ gdy } d_i^* < \bar{d}_i \end{cases} \tag{17}$$

Czasem przyjmuje się normalizację $\bar{d}_i = 0$, ale nie jest to konieczne. Celem normalizacji jest uproszczenie wyprowadzenia postaci analitycznej formy funkcyjnej.

Zmienna, której wartości obserwujemy w próbie $y_i$ powstaje jako

$$y_i = y_i^* * d_i \tag{18}$$

Zestawiając powyższe równania możemy przedstawić model nielosowego doboru w postaci strukturalnej

$$\begin{cases} y_i^* = X_i'\beta + \varepsilon_i \\ d_i^* = Z_i'\gamma + \nu_i \\ d_i = \begin{cases} 1 \text{ gdy } d_i^* \geq \bar{d}_i \\ 0 \qquad d_i^* < \bar{d}_i \end{cases} \\ y_i = y_i^* * d_i \end{cases} \tag{19}$$

Dodatkowo przyjmuje się, że $\varepsilon_i$ oraz $\nu_i$ są składnikami losowymi o średniej zero, wariancjach odpowiednio $\sigma_\varepsilon^2$ i $\sigma_\nu^2$ oraz kowariancji $cov(\varepsilon_i, \nu_i) = \sigma_{\varepsilon\nu}$. W przypadku, gdy kowariancja jest równa zero, a składniki losowe mają rozkłady normalne, równania (15) i (16) są niezależne i nie występuje problem selekcji w modelu.

Z reguły przedstawiając model prezentuje się dwa pierwsze równania. Równanie zmiennej $d_i^*$ nazywane jest równaniem selekcji, a równanie opisujące $y_i^*$ równaniem zjawiska.

Przypuśćmy, że zmienne losowe $Y$ i $Z$ mają dwuwymiarowy rozkład z korelacją $\rho$. Interesuje nas rozkład zmiennej $Y$ pod warunkiem, że zmienna $Z$ jest większa od pewnej ustalonej wartości. Intuicja nam podpowiada, że jeżeli zmienne $Y$ i $Z$ są dodatnio skorelowane to obcięcie zmiennej $Z$ powinno przesunąć dystrybuantę rozkładu zmiennej $Y$ w prawo. Jesteśmy zainteresowani zarówno rozkładem obciętej zmiennej losowej, jak również jego wartością średnią i wariancją.

Na początku dla uproszczenia analizy zajmiemy się dwuwymiarowym rozkładem normalnym. Obcięty łączny rozkład zmiennych $Y$ i $Z$ jest dany następująco:

$$f(y, z \mid Z > a) = \frac{f(y, z)}{Pr(Z > a)} \tag{20}$$

By uzyskać rozkład brzegowy przypadkowo obciętej zmiennej $Y$ wystarczy dokonać całkowania powyższego wyrażenia po zmiennej $Z$. Momenty przypadkowo obciętego rozkładu normalnego wynoszą odpowiednio:

$$E[Y \mid Z > a] = \int y \frac{f(y, z)}{Pr(Z > a)} dz = \mu_y + \rho\sigma_y\lambda(\alpha_z) \tag{21}$$

$$Var[Y \mid Z > a] = \int (Y - EY)^2 \frac{f(y,z)}{Pr(Z > a)} dz = \sigma_y^2[1 - \rho^2 \delta(\alpha_z)]$$

gdzie: $\alpha_z = (a - \mu_z)/\sigma_z$, $\lambda(\alpha_z) = \phi(\alpha_z)/[1 - \Phi(\alpha_z)]$, $\delta(\alpha_z) = \lambda(\alpha_z)[\lambda(\alpha_z) - \alpha_z]$, $\Phi(\alpha_z)$ jest dystrybuantą rozkładu normalnego, a $\phi(\alpha_z)$ jego gęstością.

Odpowiednie wyrażenia opisujące momenty zmiennej $Z$ wyglądają następująco:

$$E[z \mid \text{ obcięcie}] = \mu_z + \sigma_y \lambda(\alpha_z) \tag{22}$$

$$Var[z \mid \text{ obcięcie}] = \sigma_z^2[1 - \delta(\alpha_z)]$$

Jeżeli obcięcie rozkładu następuje z góry. tzn. $z < a$, należy dokonać podstawienia $\lambda(\alpha_z) = -\phi(\alpha_z)/\Phi(\alpha_z)$

Tak jak można oczekiwać, średnia rozkładu obciętego jest przesunięta w kierunku wyznaczonym przez znak korelacji pomiędzy zmiennymi jeżeli obcięcie następuje z dołu, a w przeciwnym kierunku do znaku korelacji przy obcięciu rozkładu z góry. Dodatkowo obcięcie rozkładu zmniejsza wariancję, ponieważ zarówno $\delta(\alpha_z)$ jak i $\rho^2$ są pomiędzy zerem a jedynką.

### 1.4.6 Przykład 5. Regresja w modelu z selekcją

W celu uzasadnienia celowości użycia modelu powstałego w wyniku obcięcia rozkładów rozważymy przykład.

1. *Równanie płac.* Różnica między rynkową pensją jednostki, której może domagać się na rynku pracy, i jej minimalną akceptowaną płacą, powyżej której osoba zdecyduje się uczestniczyć w rynku pracy jest funkcją takich charakterystyk jak wiek i wykształcenie, oraz, np. liczba dzieci i gdzie dana osoba mieszka.

2. *Równanie godzin pracy.* Pożądana liczba godzin pracy dostarczana na rynek zależy od płacy, charakterystyk gospodarstwa domowego np.: czy występują w nim małe dzieci, stanu cywilnego osoby.

Problem obcięcia występuje ponieważ równanie opisuje pożądaną liczbę godzin pracy, ale wartość ta jest obserwowana tylko w przypadku gdy badana osoba pracuje. Wnioskujemy z tego, że płaca rynkowa przekracza płacę minimalną przy której dana osoba skłonna jest podjąć pracę. Z tego powodu zmienna opisująca ilość godzin w drugim równaniu jest przypadkowo obcięta.

Występuje systematyczna różnica pomiędzy wylosowaną próbą a populacją z której ona pochodzi. By opisać powyższe przykłady za pomocą ogólnej struktury równań, możemy posłużyć się nieobserwowaną zmienną. Równanie selekcji będzie dane następująco:

$$d_i^* = Z'\gamma_i + \nu_i$$

równanie zjawiska przyjmuje następującą formę:

$$y_i = X_i'\beta + \varepsilon_i$$

Zmienna $y_i$ jest obserwowana, tylko w przypadku gdy obserwacja znajdzie się w próbie, czyli gdy $d_i^* > 0$.

Przyjmijmy dodatkowo, że $\nu_i$ i $\varepsilon_i$ mają dwuwymiarowy rozkład normalny ze średnimi zero i korelacją $\rho$. Przyjmując założenie o normalności rozkładu mamy:

$$\varepsilon_i = \rho u_i + \xi_i \tag{23}$$

gdzie $\xi$ jest niezależne od $u_i$. Do wyprowadzenia postaci funkcyjnej modelu przydaje się użyteczny fakt o rozkładzie normalnym:

$$E(\nu_i \mid \nu_i > -x) = \lambda(x) = \frac{\phi(x)}{\Phi(x)} \tag{24}$$

Funkcja $\lambda(x)$ nazywana jest odwróconym ilorazem Millsa. Jest ona iloczynem korelacji między składnikami losowymi i wariancji składnika losowego równania selekcji.

Przyjmując te założenia możemy uzyskać model, który jest dostosowany do obserwacji w naszej próbie:

$$\begin{aligned} E[y_i \mid y_i \text{ jest obserwowane}] &= E[y_i \mid d_i^* > 0] \\ &= E[y_i \mid \nu_i > -Z'\gamma_i] \\ &= X_i'\beta + E[\varepsilon_i \mid \nu_i > -Z'\gamma_i] \\ &= X_i'\beta + \rho\sigma_\varepsilon\lambda_i(\alpha_\nu) \\ &= X_i'\beta + \beta_\lambda\lambda_i(\alpha_\nu) \end{aligned}$$

gdzie: $\alpha_u = -Z_i'\gamma/\sigma_u$ oraz $\lambda_i(\alpha_u) = \phi(Z_i'\gamma/\sigma_u)/\Phi(Z_i'\gamma/\sigma_u)$

$$E[Y_i \mid d_i^* > 0] + \nu_i = X_i\beta + \beta_\lambda + \lambda_i(\alpha_\nu) + \nu_i$$

Regresja przeprowadzona metodą najmniejszych kwadratów z użyciem dostępnych obserwacji, np. regresja przepracowanych godzin na ich determinanty z użyciem wyłącznie obserwacji dla pracujących kobiet daje niezgodne estymatory wektora parametrów $\beta$. Możemy potraktować ten problem jako problem zmiennych pominiętych. Regresja metodą MNK $y$ na $X$ i $\lambda$ daje zgodne estymatory, lecz jeśli $\lambda$ jest pominięta w regresji, wtedy powstaje błąd na skutek błędnej specyfikacji modelu w wyniku pominięcia zmiennej.

$$y_i = X_i'\beta + \rho\lambda(Z_i\gamma) + \nu_i \tag{25}$$

Ta regresja oszacowana MNK daje zgodne estymatory dla obserwowanej podpróby. Z drugiej części równania (22) wynika, że jeśli nawet $\lambda_i$ są obserwowane, metoda najmniejszych kwadratów będzie dawała nieefektywne estymatory, ponieważ zaburzenie losowe $\nu_i$ jest heteroscedastyczne.

Krańcowy efekt wpływu regresora na zmienną objaśnianą $y_i$ w obserwowanej próbie zawiera dwa składniki. $\beta$ jest bezpośrednim efektem mierzonym na poziomie średniej wartości $Y_i$. Dodatkowo, dla niektórych zmiennych niezależnych, jeśli $Z_i^* > 0$ wtedy oddziaływuje na zmienną $y_i$ poprzez parametr $\lambda_i$. Całkowity efekt regresora, który występuje zarówno w $X_i$ jak i w $Z_i$ na zmienną $y$ wynosi:

$$\frac{\partial E[Y_i \mid d_i^* > 0]}{\partial X_{i,k}} = \beta_k - \gamma_k(\frac{\rho\sigma_\epsilon}{\sigma_u})\delta_i(\alpha_u) \tag{26}$$

gdzie:

$$\delta_i = \lambda_i^2 - \alpha_i\lambda_i$$

Przypuśćmy, że $\rho$ jest większe od zera oraz $E[Y_i]$ jest większe przy dodatnim $d_i^*$ niż przy ujemnym. Ponieważ $\delta_i$ należy do przedziału (0,1), dodatkowy składnik w równaniu (26) redukuje efekt krańcowy. Zmiana prawdopodobieństwa wpływa na średnią wartość $Y_i$ - w grupie z dodatnim $d_i^*$ średnia wartość jest wyższa. Drugi składnik równania kompensuje ten efekt. W rezultacie otrzymujemy warunkową średnią zmianę.

### 1.4.7 Przykład 6.

Przypuśćmy, że wykształcenie wpływa na prawdopodobieństwo migracji i dochód. Jeżeli założymy, że dochody migrantów są wyższe niż osób o takich samych charakterystykach ale nie migrujących, wtedy krańcowy efekt wykształcenia składa się z dwóch części: pierwsza powstaje w wyniku zwiększenia prawdopodobieństwa awansu do grupy o wyższych dochodach, druga w wyniku wpływu na dochód wewnątrz danej grupy dochodowej. Współczynnik przy zmiennej wykształcenie przeszacowuje krańcowy efekt wykształcenia ludzi migrujących i niedoszacowuje ten efekt dla pozostałych. Oszacowane kierunki i rozmiary efektów mogą różnić się od otrzymanych z wektora $\beta$.

W większości przypadków, zmienna decydująca o wyborze do próby $d^*$ nie jest obserwowana. Często znamy jedynie jej znak. Np. z reguły wiemy czy kobieta pracuje czy nie, lub czy osoba migrowała czy nie. Z takiej informacji możemy uzyskać znak zmiennej $d^*$, ale nic nie wiemy o jej wartości. Ponieważ skala zmiennej $d^*$ nie jest znana, nie może być oszacowana wariancja składnika losowego. Dlatego przy szacowaniu modelu za pomocą procedury dwustopniowej jej wartość arbitralnie jest ustalana na 1.

### 1.4.8 Estymacja

Parametry modelu z selekcjonującą się próbą mogą być estymowane metodą największej wiarogodności. Ta procedura w literaturze anglojęzycznej jest określana jako *Full Maximum Likelihood*. Wymaga przyjęcia trudnego do spełnienia założenia o dwuwymiarowej normalności rozkładu $\nu_i$ i $\varepsilon_i$. Wówczas łącznie są znajdowane oszacowania parametrów $\beta, \gamma, \rho$.

Ale, zazwyczaj używana jest dwustopniowa procedura Heckmana. Polega ona na tym, że:

1. Estymowany jest model probitowy metodą największej wiarogodności. Z tego modelu uzyskuje się estymator $\gamma$. Następnie dla każdej obserwacji wylicza się $\hat{\lambda}_i = \phi(X_i'\hat{\gamma})/\Phi(X_i'\hat{\gamma})$, oraz $\hat{\delta}_i = \hat{\lambda}_i(\hat{\lambda}_i - w_i'\hat{\gamma})$.

2. Przeprowadzana jest regresja metodą MNK $y$ na $X$ i $\hat{\lambda}$ i z niej otrzymujemy $\beta$ oraz $\beta_\lambda = \rho\sigma_\varepsilon$.

Możliwe jest również uzyskanie zgodnych estymatorów dla indywidualnych parametrów $\rho$ i $\sigma_\varepsilon$. Dla każdej obserwacji, prawdziwa warunkowa wariancja wynosi:

$$\sigma_i^2 = \sigma_\varepsilon^2(1 - \rho^2\delta_i)$$

Średnia warunkowa wariancja w próbie dąży do

$$p\lim \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = \sigma_\varepsilon^2(1 - \rho^2\bar{\delta})$$

czyli do wariancji otrzymywanej metodą najmniejszych kwadratów $\frac{e'e}{n}$. Dla kwadrat współczynnika $\lambda$ mamy

$$p\lim b_\lambda^2 = \rho^2\sigma_\epsilon^2$$

Z modelu probitowego otrzymujemy:

$$p\lim \sum_{i=1}^{n}\hat{\delta}_i = \bar{\delta}.$$

Możemy uzyskać zgodny estymator parametru $\sigma_\varepsilon^2$ używając:

$$\sigma_\varepsilon^2 = \frac{1}{n}e'e + \hat{\bar{\delta}}b_\lambda^2$$

ostatnim estymatorem jest oszacowanie korelacji składników losowych:

$$\hat{\rho^2} = \frac{b_\lambda^2}{\hat{\sigma}_\epsilon^2}$$

który uzupełnia zbiór estymatorów modelu.

Ten sposób estymacji jest często używany przy występowaniu problemu samoselekcji, bądź braku losowości próby. Jednak stosowanie procedury Heckmana wymaga ostrożności bowiem jest ona wrażliwa na niespełnienie założenia o normalności rozkładu reszt. często dodatkowym utrudnieniem jest mała wariancja parametru $\lambda(w_i\gamma)$ w próbie. Jest ona konsekwencją słabego dopasowania modelu probitowego do danych empirycznych.

### 1.4.9 Przykład 7.

Dane do przykładu sa wycinkiem amerykańskiego badania rynku pracy. Próba zawiera informacje o 2000 kobiet, wśród których 657 nie pracuje. Zmienna `lw` to logarytm płacy, `age` to wiek w latach, `children` oznacza ilość posiadanych dzieci, a `education` ilość ukończonych lat nauki.

W celu identyfikacji parametrów przyjmujemy że stan cywilny wpływa na partycypację w rynku pracy, ale nie wpływa na wysokość płacy.

```
. use http://www.stata-press.com/data/imeus/womenwk, clear

. heckman lw education age children, select(age married children education)

Iteration 0:   log likelihood = -1065.7948
Iteration 1:   log likelihood = -1053.6855
Iteration 2:   log likelihood =  -1052.867
Iteration 3:   log likelihood = -1052.8574
Iteration 4:   log likelihood = -1052.8574

Heckman selection model                         Number of obs      =      2000
(regression model with sample selection)        Censored obs       =       657
                                                Uncensored obs     =      1343

                                                Wald chi2(3)       =    454.78
Log likelihood = -1052.857                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lw           |
   education |   .0397189   .0024525    16.20   0.000     .0349121    .0445256
         age |   .0075872   .0009748     7.78   0.000     .0056767    .0094977
    children |  -.0180477   .0064544    -2.80   0.005    -.0306981   -.0053973
       _cons |   2.305499   .0653024    35.30   0.000     2.177509     2.43349
-------------+----------------------------------------------------------------
select       |
         age |   .0350233   .0042344     8.27   0.000     .0267241    .0433225
     married |   .4547724   .0735876     6.18   0.000     .3105434    .5990014
```

```
    children |  .4538372   .0288398    15.74   0.000     .3973122    .5103621
   education |  .0565136   .0110025     5.14   0.000     .0349492    .0780781
       _cons | -2.478055   .1927823   -12.85   0.000    -2.855901   -2.100208
-------------+----------------------------------------------------------------
     /athrho |  .3377674   .1152251     2.93   0.003     .1119304    .5636045
    /lnsigma | -1.375543   .0246873   -55.72   0.000    -1.423929   -1.327156
-------------+----------------------------------------------------------------
         rho |  .3254828   .1030183                      .1114653    .5106469
       sigma |  .2527024   .0062385                      .2407662    .2652304
      lambda |  .0822503   .0273475                      .0286501    .1358505
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =     5.53   Prob > chi2 = 0.0187
------------------------------------------------------------------------------
```

Test ilorazu wiarogodności odrzuca hipotezę o niezależności równania selekcji i równania zjawiska. Oznacza to, że pominięcie mechanizmu selekcji i oszacowanie modelu regresji prowadziłoby do uzyskania obciążonych estymatorów.

Oszacowania uzyskane metodą największej wiarogodności są czułe na niespełnienie założenia o dwuwymiarowej normalności. Dlatego często używa się procedury dwustopniowej, która jest bardziej odporna.

```
. heckman lw education age children,select(age married children education) twostep

Heckman selection model -- two-step estimates   Number of obs      =      2000
(regression model with sample selection)        Censored obs       =       657
                                                Uncensored obs     =      1343

                                                Wald chi2(6)       =    737.21
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lw           |
   education |  .0427067    .003106    13.75   0.000     .0366191    .0487944
         age |   .009322   .0014343     6.50   0.000     .0065108    .0121333
    children | -.0019549   .0115202    -0.17   0.865    -.0245341    .0206242
       _cons |  2.124787   .1249789    17.00   0.000     1.879833    2.369741
-------------+----------------------------------------------------------------
select       |
         age |  .0347211   .0042293     8.21   0.000     .0264318    .0430105
     married |  .4308575    .074208     5.81   0.000     .2854125    .5763025
    children |  .4473249   .0287417    15.56   0.000     .3909922    .5036576
   education |  .0583645   .0109742     5.32   0.000     .0368555    .0798735
       _cons | -2.467365   .1925635   -12.81   0.000    -2.844782   -2.089948
-------------+-----------------------------------------------------------------
mills        |
```

```
      lambda |  .1822815   .0638285     2.86   0.004       .05718     .307383
-------------+----------------------------------------------------------------
         rho |   0.66698
       sigma | .27329216
      lambda | .18228151   .0638285
------------------------------------------------------------------------------
```

Dzieci zwiększają prawdopodobieństwo posiadania pracy, ale zmiejszają płacę.

Jak widać wielkości oszacowań parametrów wektorów $\beta$ oraz $\gamma$ nieznacznie różnią się. Natomiast występuje znaczna różnica w oszacowanej wielkości parametru $\rho$. Może to wskazywać na brak dwuwymiarowej normalności rozkładu składnika losowego, bowiem metoda dwustopniowa daje oszacowania zgodne i bardziej odporne.

# Literatura

[1] Takeshi Amemya (1981) „Qualitative Resopnse Models:A Survey” *Journal of Economic Literature*, vol 19 pp. 1483-1536.

[2] Kit Baum (2006) *An Introduction to Econometrics Using Stata*, Stata Press.

[3] William H. Greene (2003) *Econometric Analysis*, 5th edition.

[4] J. Scott Long, Jeremy Freese (2003) *Regression Models for Categorical Dependent Variables Using Stata. Revised Edition*, Stata Press.

[5] Thomas A. Mroz (1987) „The Sensivity of an Empirical Model of Marired Women’s Hours of Worked to Economic and statistical Assumptions” *Econometrica*, vol. 55/4 pp. 765-799.

[6] Wojciech Niemiro (1999) *Rachunek Prawdopodobieństwa i Statystyka Matematyczna*, Szkoła Nauk Ścisłych.

[7] Robert Willis, Shervin Rosen (1979) „Education and Self Selection” *Journal of Political Economy*, vol. 87/5 pp. S7-S36.