Binarne zmienne zależne

- Zmienna zależna nie jest ciągła ale przyjmuje zero lub jeden
- **Przykład:** szukanie determinant bezrobocia (próba przekrojowa)
  - zmienna objaśniana: zerojedynkowa (pracujący, bezrobotny)
  - szukamy związku między charakterystykami ekonomiczno-społecznymi a byciem bezrobotnym
- W modelach dla zmiennych binarnych staramy się wyjaśnić prawdopodobieństwa stanów za pomocą zmiennych niezależnych
- Prawdopodobieństwo pojedynczego zdarzenia dane jest rozkładem

dwumianowym

$$\Pr\left(y_i \middle| \boldsymbol{x}_i\right) = \begin{cases} 1 - p\left(\boldsymbol{x}_i\right) & \text{dla} \quad y_i = 0 \\ p\left(\boldsymbol{x}_i\right) & \text{dla} \quad y_i = 1 \end{cases}$$

gdzie $p\left(\boldsymbol{x}_i\right) = F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$ i $F\left(\cdot\right)$ jest zazwyczaj dystrybuantą jakiegoś rozkładu prawdopodobieństwa.

- Estymator parametru $\boldsymbol{\beta}$ liczymy za pomocą MNW.

Interpretacja ekonomiczna

Podejście indeksowe

- y^* jest użytecznością lub zyskiem netto
- y^* jest nieobserwowalne
- y^* jest losowe

  $$y^* = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

  gdzie ε ma symetryczną dystrybuantę z $\mathrm{E}\,(\boldsymbol{\varepsilon}) = 0$ i $\mathrm{Var}\,(\boldsymbol{\varepsilon}) = 1$
- Parametry modelu szacujemy na postawie obserwowalnych wybów y, na które zdeterminowane są nieobserwowalnym y^*.

- Wybór $y = 1$ jest podejmowany jeśli przyrost związanej z nim użyteczności (zysku) jest dodatni

$$\begin{array}{lll} y = 0 & \text{dla} & y^* \leq 0 \\ y = 1 & \text{dla} & y^* > 0 \end{array}$$

- Prawdopodobieństwo, że $y = 1$ jest równe

$$\begin{aligned} \Pr\{y = 1\} &= \Pr\{y^* > 0\} = \Pr\{\boldsymbol{x\beta} + \boldsymbol{\varepsilon} > 0\} \\ &= \Pr\{\boldsymbol{\varepsilon} > -\boldsymbol{x\beta}\} \end{aligned}$$

- Skoro dystybuanta $\boldsymbol{\varepsilon}$ jest symetryczna, to

$$\Pr\{y = 1\} = \Pr\{\boldsymbol{\varepsilon} < \boldsymbol{x\beta}\} = F(\boldsymbol{x\beta})$$

gdzie $F(\cdot)$ jest dystrybuantą $\boldsymbol{\varepsilon}$

- W tym podejściu, wybór zależy jedynie od charakterystyk podmiotu i alternatywy wybranej

Losowa użyteczność

- Załóżmy, że y_0^* i y_1^* są użytecznościami z działań 0 i 1.

- y_0^* i y_1^* są nieobserwowalne i dane równaniami

$$y_0^* = \boldsymbol{x}_0\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0$$
$$y_1^* = \boldsymbol{x}_1\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_1$$

- Załóżmy, że $\boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}_1$ mają symetryczną dystrybuantę

- Obserwujemy y takie że

$$\begin{array}{lll} y = 0 & \text{if} & y_0^* > y_1^* \\ y = 1 & \text{if} & y_0^* \leq y_1^* \end{array}$$

- Podmiot podejmuje działanie 0 jeśli y_0^* jest lepsze niż y_1^* i 1 jeśli y_1^* jest lepsze niż y_0^*.

- Prawdopodobieństwo zaobserwowania 1 jest równe:

$$\begin{aligned} \Pr\{y = 1\} &= \Pr\{y_0^* \leq y_1^*\} = \Pr\{\boldsymbol{x}_0\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0 \leq \boldsymbol{x}_1\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_1\} \\ &= \Pr\{\boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_0 \geq \boldsymbol{x}_0\boldsymbol{\beta} - \boldsymbol{x}_1\boldsymbol{\gamma}\} = F(\boldsymbol{x}_1\boldsymbol{\gamma} - \boldsymbol{x}_0\boldsymbol{\beta}) \end{aligned}$$

- W modelu losowej użyteczności wybór zależy nie tylko od charakterystyk podmiotu, charakterystyk alternatywy wybranej, ale też od charakterystyk alternatywy niewybranej.

Interpretacja współczynników

- Wartość oczekiwana zmiennej zależnej w modelu dla zmiennej binarnej jest równa:

$$\mathrm{E}\left(y_i \middle| \boldsymbol{x}_i\right) = 0\left[1 - F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right] + 1F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = p\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$$

- Ponieważ przyjęta jest konwencja, że 1 oznacza sukces a 0 porażkę więc, wartość oczekiwana zmiennej zależnej w modelu dla zmiennej binarnej równa jest prawdopodobieństwu sukcesu.

- Efekt cząstkowy jest równy:

$$\frac{\partial \mathrm{E}\left(y \middle| \boldsymbol{x}\right)}{\partial \boldsymbol{x}} = \frac{\partial p\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)}{\partial \boldsymbol{x}} = f\left(\boldsymbol{x}\boldsymbol{\beta}\right)\boldsymbol{\beta}$$

- Efekt cząstkowy w tym przypadku intepretujemy jako wpływ jednostkowej zmiany zmiennej niezależnej na prawdopodobieństwo sukcesu.

- Ze wzoru widać, że wielkość efektu cząstkowego zależy od tego w jakim punkcie jest on liczony

- Zazwyczaj liczymy go dla średnich wartości zmiennych egzogenicznych.

- Dla zero-jedynkowych zmiennych objaśniających efekt krańcowy definiuje się jako $\Delta F\left(\overline{\boldsymbol{x}}\boldsymbol{\beta}\right) = F\left(\overline{\boldsymbol{x}}_1\boldsymbol{\beta}\right) - F\left(\overline{\boldsymbol{x}}_0\boldsymbol{\beta}\right)$ - a więc różnice między prawdopodobiestwem sukcesu dla zmiennej zerojedynkowej równej zeru i równej 1 przy wszystkich pozostałych zmiennych ustalonych na poziomie średnich.

- Ponieważ funkcja gęstości $f\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$ jest zawsze dodatnia więc znak efektu cząskowego jest równy znakowi współczynnika przy zmiennej. Wynika z

tego, że jakkolwiek same współczynniki nie są zazwyczaj interpretowalne to można intepretować ich znaki.

- Można jednak intepretować proporcje między współczynnkimi ponieważ są one równe proporcjom między efektami krańcowymi

$$\frac{\partial \mathrm{E}\left(y|\,\boldsymbol{x}\right)}{\partial x_j} \Big/ \frac{\partial \mathrm{E}\left(y|\,\boldsymbol{x}\right)}{\partial x_k} = \frac{\boldsymbol{\beta}_j}{\boldsymbol{\beta}_k}$$

Miary dopasowania

- Często stosowaną miarą dopasowania jest

$$pseudo\text{-}R^2 = 1 - \frac{\ell\left(\widetilde{\boldsymbol{\beta}}\right)}{\ell\left(\widetilde{\boldsymbol{\beta}}_R\right)}$$

  gdzie $\widetilde{\boldsymbol{\beta}}_R$ zawiera jedynie stałą

- $pseudo\text{-}R^2$ spełnia

$$0 \leq pseudo\text{-}R^2 \leq 1$$

- $pseudo\text{-}R^2$ nie można jednak w pełni ściśle interpretować jako procent zmienności wyjaśnionej przez model.

- Innym sposobem mierzenia jakości dopasowania jest analiza tablicy częstości:

Prognozowane \ Zaobserwowane	0	1	Razem
0	n_{00}	n_{01}	$n_{00}+n_{01}$
1	n_{10}	n_{11}	$n_{10}+n_{11}$
Razem	$n_{00}+n_{10}$	$n_{01}+n_{11}$	

- Arbitralnie zakładamy, że model przewiduje $y_i = 1$ jeśli $\widehat{p}_i = \boldsymbol{x}_i\boldsymbol{b} > 0.5$

Liniowy model prawdopodobieństwa (LPM)

- W liniowym modelu prawdopodobieństwa LPM (**L**inear **P**robability **M**odel) zakładmy, że dystrybuanta rozkładu ma postać funkcji liniowej:

$$p\left(\boldsymbol{x}_i\right) = F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = \boldsymbol{x}_i\boldsymbol{\beta}$$

- W tym przypadku estymacja parametrów $\boldsymbol{\beta}$ sprawadza się do szacowania MNK równania

$$y_i = \boldsymbol{x}_i\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

- Podejście to mimo swojej prostoty związane jest z dwoma poważnymi problemami:

1. Błędy losowe w równaniu regresji będą heteroskedastyczne. Zmienna ma rozkład dwumianowy więc:

$$\mathrm{E}\left(y_i\right) = F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = \boldsymbol{x}_i\boldsymbol{\beta}$$

$$\mathrm{E}\left(y_i^2\right) = 0^2\left(1 - \boldsymbol{x}_i\boldsymbol{\beta}\right) + 1^2\boldsymbol{x}_i\boldsymbol{\beta} = \boldsymbol{x}_i\boldsymbol{\beta}$$

$$\begin{aligned}\mathrm{Var}\left(y_i\right) &= \mathrm{E}\left(x_i^2\right) - \mathrm{E}\left(x_i\right)^2 = \boldsymbol{x}_i\,\boldsymbol{\beta} - \left(\boldsymbol{x}_i\boldsymbol{\beta}\right)^2 \\ &= \boldsymbol{x}_i\boldsymbol{\beta}\left(1 - \boldsymbol{x}_i\boldsymbol{\beta}\right)\end{aligned}$$

   Wariancja y_i zależy od $\boldsymbol{x}_i$.
2. Nie ma gwarancji, że $\widehat{y}_i = \boldsymbol{x}_i\ \boldsymbol{b} \in [0,1]$. Ponieważ $\widehat{y}_i$ interpretujemy jako prawdopodobieństwa stanów - ujemne lub większe od 1 wartości dopasowane będą nieinterpretowalne

- Mimo tych wad liniowy model prawdopodobieństwa jest dosyć często

używany w praktycznych zastosowaniach ze względu na łatwość estymacji.

- Efekty cząskowe w tym modelu równe są poprostu $\boldsymbol{\beta}$
- Wpływu heteroskedastyczności na wyniki wnioskowania można wyeliminować stosując odporną macierz wariancji White'a
- Hetroskedastyczność można też wyeliminować za pomocą $UMNK$:
  - estymujemy LPM za pomocą MNK
  - liczymy wariancje $\widehat{\boldsymbol{\sigma}}_i^2 = \boldsymbol{x}_i\boldsymbol{b}\left[1-\boldsymbol{x}_i\boldsymbol{b}\right] = \widehat{y}_i\left(1-\widehat{y}_i\right)$ (wór ten ma sens jedynie dla $\widehat{y}_i \in [0,1]$)
  - liczymy regresję $y_i/\widehat{\boldsymbol{\sigma}}_i$ na $x_{1i}/\widehat{\boldsymbol{\sigma}}_i, \ldots, x_{Ki}/\widehat{\boldsymbol{\sigma}}_i$

Przykład Zależność aktywności zawodowej od wieku i wykształcenia: dane BAEL - pytanie: czy w ostatnim tygodniu wykonywał Pan(i) jakąkolwiek pracę?

```
      Source |       SS       df       MS              Number of obs =   47860
-------------+------------------------------           F(  9, 47850) =  993.61
       Model |  1827.14253     9  203.015837            Prob > F      =  0.0000
    Residual |  9776.79083 47850  .204321648            R-squared     =  0.1575
-------------+------------------------------           Adj R-squared =  0.1573
       Total |  11603.9334 47859  .242460841            Root MSE      =  .45202

------------------------------------------------------------------------------
   _Iworks_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |  -.0036288   .0001188   -30.55   0.000    -.0038616   -.0033961
   _Iwyksz_2 |  -.1155233   .0141975    -8.14   0.000    -.1433507    -.087696
   _Iwyksz_3 |  -.1575525   .0082029   -19.21   0.000    -.1736303   -.1414747
   _Iwyksz_4 |  -.3616109   .0098378   -36.76   0.000     -.380893   -.3423287
   _Iwyksz_5 |   -.196475   .0078076   -25.16   0.000    -.2117779    -.181172
  _Iwyksz_60 |  -.7496461   .0163531   -45.84   0.000    -.7816984   -.7175937
  _Iwyksz_61 |  -.4814629   .0077523   -62.11   0.000    -.4966576   -.4662683
  _Iwyksz_70 |  -.5166902   .0162939   -31.71   0.000    -.5486266   -.4847539
```

```
  _Iwyksz_71 |   -.594044   .0302346   -19.65   0.000    -.6533043   -.5347837
       _cons |   .8569807   .0084361   101.59   0.000     .8404459    .8735155
------------------------------------------------------------------------------
```

- Wyniki testu na heteroskedastyczność:

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of _Iworks_1

         chi2(1)      =   961.62
         Prob > chi2  =   0.0000
```

- Niektóre wartości przewidywane:

```
     +-----------+
     |      yhat |
     |-----------|
  1. | -.1974881 |
  2. | -.1793439 |
  3. | -.1720862 |
```

```
  4. | -.1466843 |
  5. | -.1398648 |
```

- Jednak z 47860 obserwacji tylko $144 \notin (0,1)$

Probit

- Dystrybuanta $F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$, z rozkładu normalnego.
- Funkcja prawdopodobieństwa zajścia pojedynczego zdarzenia:

$$\Pr\left(y_i\right) = \begin{cases} 1 - \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) & \text{dla} \quad y_i = 0 \\ \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) & \text{dla} \quad y_i = 1 \end{cases}$$

$$= \left[1 - \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right]^{1-y_i} \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)^{y_i}$$

- Funkcja wiarygodności ma postać

$$L\left(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\right) = \prod_{i=1}^{n} \left(1 - \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right)^{1-y_i} \Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)^{y_i}$$

- Logarytm funkcji wiarygodności będzie ma postać

$$\ell\left(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\right)=\sum_{i=1}^{n}\left[\left(1-y_i\right)\ln\left(1-\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right)+y_i\ln\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right]$$

- Często podaje się jedynie $\ell_i\left(\boldsymbol{\beta}\right)$

$$\ell_i\left(\boldsymbol{\beta}\right)=\left(1-y_i\right)\ln\left(1-\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right)+y_i\ln\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}\right).$$

- Efekt cząstkowy

$$\frac{\partial\,\mathrm{E}\left(y|\,\boldsymbol{x}\right)}{\partial\boldsymbol{x}}=\boldsymbol{\phi}\left(\boldsymbol{x}\boldsymbol{\beta}\right)\boldsymbol{\beta}$$

Przykład Zależność aktywności zawodowej od wieku i wykształcenia: kontynuacja

- Współczynniki:

```
Probit estimates                                  Number of obs   =      47860
                                                  LR chi2(9)      =    8274.24
                                                  Prob > chi2     =     0.0000
Log likelihood = -28311.095                       Pseudo R2       =     0.1275

------------------------------------------------------------------------------
   _Iworks_1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |  -.0113366   .0003666   -30.93   0.000     -.012055   -.0106181
   _Iwyksz_2 |  -.3254287   .0404874    -8.04   0.000    -.4047825   -.2460749
   _Iwyksz_3 |   -.433603    .023857   -18.18   0.000     -.480362   -.3868441
   _Iwyksz_4 |  -.9675114   .0286116   -33.82   0.000    -1.023589   -.9114338
   _Iwyksz_5 |  -.5330219   .0227775   -23.40   0.000    -.5776649   -.4883789
  _Iwyksz_60 |  -2.513776   .0730264   -34.42   0.000    -2.656905   -2.370647
  _Iwyksz_61 |  -1.343109   .0233078   -57.62   0.000    -1.388792   -1.297427
  _Iwyksz_70 |  -1.678283   .0650819   -25.79   0.000    -1.805842   -1.550725
  _Iwyksz_71 |  -2.321871   .1804895   -12.86   0.000    -2.675624   -1.968118
```

```
       _cons |   1.027194   .0256271    40.08   0.000     .9769656    1.077422
------------------------------------------------------------------------------
```

- Efekty krańcowe:

```
Probit estimates                                  Number of obs =  47860
                                                  LR chi2(9)    =8274.24
                                                  Prob > chi2   = 0.0000
Log likelihood = -28311.095                       Pseudo R2     = 0.1275

------------------------------------------------------------------------------
_Iwork~1 |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
    wiek |  -.0043578   .0001406   -30.93   0.000   43.9481  -.004633 -.004082
_Iwyks~2*|  -.1181235   .0136421    -8.04   0.000   .027267  -.144862 -.091385
_Iwyks~3*|  -.1582424   .0081461   -18.18   0.000    .19135  -.174209 -.142276
_Iwyks~4*|  -.3006814   .0063921   -33.82   0.000   .082804   -.31321 -.288153
_Iwyks~5*|  -.1946251   .0077792   -23.40   0.000    .26728  -.209872 -.179378
_Iwyk~60*|  -.4090495   .0025428   -34.42   0.000   .020079  -.414033 -.404066
_Iwyk~61*|  -.4366777   .0058916   -57.62   0.000   .290639  -.448225  -.42513
_Iwyk~70*|  -.3783463   .0048507   -25.79   0.000   .020664  -.387854 -.368839
_Iwyk~71*|  -.3921597   .0035269   -12.86   0.000   .004952  -.399072 -.385247
```

```
---------+--------------------------------------------------------------------
  obs. P |   .4131425
 pred. P |   .3926327  (at x-bar)
---------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| are the test of the underlying coefficient being 0
```

Logit

- Model logitowy otrzymujemy dla

$$F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}} = \Lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$$

- Funkcja prawdopodobieństwa dla pojedynczej obserwacji

$$\Pr\left(y_i\right) = \begin{cases} \frac{1}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}} & \text{dla} \quad y_i = 0 \\ \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}} & \text{dla} \quad y_i = 1 \end{cases}$$

- Logarytm funkcji wiarygodności

$$\begin{aligned}
&\ell\left(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\right) \\
&= \sum_{i=1}^{n}\left[(1-y_i)\ln\left(\frac{1}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}\right)+y_i\ln\left(\frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}\right)\right] \\
&= \sum_{i=1}^{n} y_i\boldsymbol{x}_i\boldsymbol{\beta}-\sum_{i=1}^{n}\ln\left(1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}\right)
\end{aligned}$$

- Efekt cząstkowy:

$$\frac{\partial \mathrm{E}\left(y|\,\boldsymbol{x}\right)}{\partial \boldsymbol{x}}=\frac{\partial \Lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)}{\partial \boldsymbol{x}}=\Lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\left[1-\Lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right]\boldsymbol{\beta}$$

- W przypadku modelu logitowego można także bezośrednio intepretować same parametry.

- Zdefiniujmy szansę sukcesu jako stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki:

$$Odds\left(\boldsymbol{x}_i\right)=\frac{\Pr\left(y_i=1\middle|\,\boldsymbol{x}_i\right)}{\Pr\left(y_i=0\middle|\,\boldsymbol{x}_i\right)}=\frac{\frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}}{\frac{1}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}}=e^{\boldsymbol{x}_i\boldsymbol{\beta}}$$

Przykład Jeśli prawdopodobieństwo zdania egzaminu z ekonometri jeśli się w ogóle nie uczyłem $(x_i=0)$ wynosi $\frac{1}{9}$ a szansa zdania egzaminu jest równa $Odds\left(0\right)=\frac{\frac{1}{9}}{1-\frac{1}{9}}=\frac{1}{8}$

- Policzmy teraz iloraz, szansy dla zmiennej niezależnej równej $\boldsymbol{x}_i$ i szansy dla zmiennej niezależnej równej $\boldsymbol{x}_j$:

$$\frac{Odds\left(\boldsymbol{x}_i\right)}{Odds\left(\boldsymbol{x}_j\right)}=\exp\left[\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)\boldsymbol{\beta}\right]=\exp\left(\Delta\boldsymbol{x}\boldsymbol{\beta}\right)$$

Przykład Jeśli prawdopodobieństwo zdania egzaminu z ekonometri jeśli się w ogóle nie uczyłem jeden dzień $(x_i = 1)$ wynosi $\frac{1}{3}$ to szansa zdania egzaminu w tym przypadku będzie równa $Odds\,(1) = \frac{\frac{1}{3}}{1-\frac{1}{3}} = \frac{1}{2}$ a iloraz szans będzie równy

$$\frac{Odds\,(1)}{Odds\,(0)} = \frac{\frac{1}{2}}{\frac{1}{8}} = 4$$

Prawdopodbieństwo zdania egzaminu wzrosło więc 3 razy a szansa zdania egzaminu 4 razy.

- Ilorazy szans równe $\exp(\boldsymbol{\beta}_k)$ intepretuje się w sposób następujący:
  - Dla zmiennej ciągłej jest to współczynnik, przez który należy przemnożyć szansę sukcesu dla x_{ik}, by otrzymać szansę sukcesu dla $x_{ik} + 1$.
  - Dla zmiennej zerojedynkowej x_{ik} jest to współczynnik, przez który należy przemnożyć szansę sukcesu dla $x_{ik} = 0$, by otrzymać szansę

sukcesu dla $x_{ik} = 1$.

- Dla małych $\boldsymbol{\beta}_k$ wielkości $\boldsymbol{\beta}_k$ można także zintepretować, korzystając z tego, że $e^{\Delta \boldsymbol{x}\boldsymbol{\beta}} \approx 1 + \Delta \boldsymbol{x}\boldsymbol{\beta}$:
  - dla zmiennych ciagłych - ile procent zmieni się szansa sukcesu jeśli x_k zmieni się o 1
  - dla zmiennych zero jedynkowych - ile procent zmieni się szansa sukcesu jeżeli zmienna x_k przeskoczy z 0 na 1.

Przykład Zależność aktywności zawodowej od wieku i wykształcenia: kontynuacja

- Współczynniki

```
Logit estimates                                   Number of obs   =      47860
                                                  LR chi2(9)      =    8275.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -28310.607                       Pseudo R2       =     0.1275

------------------------------------------------------------------------------
   _Iworks_1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |  -.0186039   .0006038   -30.81   0.000    -.0197873   -.0174204
   _Iwyksz_2 |  -.5221753   .0655267    -7.97   0.000    -.6506053   -.3937453
   _Iwyksz_3 |  -.6942371   .0389682   -17.82   0.000    -.7706134   -.6178608
   _Iwyksz_4 |  -1.561855   .0471004   -33.16   0.000     -1.65417    -1.46954
   _Iwyksz_5 |  -.8542947   .0372593   -22.93   0.000    -.9273216   -.7812678
  _Iwyksz_60 |  -4.372065   .1570768   -27.83   0.000     -4.67993     -4.0642
  _Iwyksz_61 |  -2.207483   .0391494   -56.39   0.000    -2.284215   -2.130752
  _Iwyksz_70 |   -2.88925   .1303508   -22.17   0.000    -3.144732   -2.633767
  _Iwyksz_71 |  -4.194055    .416003   -10.08   0.000    -5.009406   -3.378704
```

```
       _cons |   1.666699   .0423531    39.35   0.000     1.583688    1.749709
------------------------------------------------------------------------------
```

- Krańcowe efekty

```
Marginal effects after logistic
      y  = Pr(_Iworks_1) (predict)
         =  .38568338
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    wiek |  -.0044078      .00014  -30.94   0.000  -.004687 -.004129   43.9481
_Iwyks~2*|  -.1148651      .06553   -1.75   0.080  -.243295  .013565   .027267
_Iwyks~3*|  -.1539078      .03897   -3.95   0.000  -.230284 -.077532    .19135
_Iwyks~4*|  -.2864085       .0471   -6.08   0.000  -.378724 -.194093   .082804
_Iwyks~5*|   -.189642      .03726   -5.09   0.000  -.262669 -.116615    .26728
_Iwyk~60*|   -.398101      .15708   -2.53   0.011  -.705966 -.090236   .020079
_Iwyk~61*|  -.4279628      .03915  -10.93   0.000  -.504694 -.351231   .290639
_Iwyk~70*|  -.3641805      .13035   -2.79   0.005  -.619663 -.108698   .020664
_Iwyk~71*|  -.3810388        .416   -0.92   0.360  -1.19639  .434312   .004952
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

- Interpetacja efektów cząstkowych:

- wzrost wieku o 1 rok zmniejsza prawdopodobieństwo posiadania pracy o 0.4% w stosunku do prawdopodobieństwa dla średnich w próbie
- osby z wykształceniem policealnym mają o 11% mniejsze prawdopodbieństwo posiadania pracy niż osoby z wykształceniem wyższym (poziom bazowy).

- Ilorazy szans

```
Logistic regression                                     Number of obs   =      47860
                                                        LR chi2(9)      =    8275.22
                                                        Prob > chi2     =     0.0000
Log likelihood = -28310.607                             Pseudo R2       =     0.1275

------------------------------------------------------------------------------
   _Iworks_1 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |   .9815681   .0005927   -30.81   0.000     .9804071    .9827304
   _Iwyksz_2 |   .5932287   .0388723    -7.97   0.000     .5217299    .6745259
   _Iwyksz_3 |   .4994553   .0194629   -17.82   0.000     .4627292    .5390964
   _Iwyksz_4 |   .2097466   .0098792   -33.16   0.000     .1912507    .2300313
   _Iwyksz_5 |   .4255833   .0158569   -22.93   0.000     .3956119    .4578252
  _Iwyksz_60 |   .0126251   .0019831   -27.83   0.000     .0092797    .0171767
  _Iwyksz_61 |   .1099771   .0043055   -56.39   0.000      .101854     .118748
  _Iwyksz_70 |   .0556179   .0072498   -22.17   0.000     .0430784    .0718075
  _Iwyksz_71 |    .015085   .0062754   -10.08   0.000     .0066749    .0340916
------------------------------------------------------------------------------
```

- Interpetacja współczynników:

– szansę posiadania pracy trzeba pomnożyć przez 0.98 aby uzyskać szansę posiadania pracy przy wieku większym o 1 - spadek o 2%
– szanse posiadania pracy dla wykształcenia wyższego (poziom bazowy) należy pomnożyć przez 0.58 aby uzyskać szansę posiadania pracy przy wykształceniu policealnym (poziom 1) - spadek o 42%

- Sukcesy i porażki przewidywane i zaobserwowane

```
  Logistic model for _Iworks_1

              -------- True --------
Classified |         D             ~D  |      Total
-----------+---------------------------+-----------
     +     |     12568           7211  |      19779
     -     |      7205          20876  |      28081
-----------+---------------------------+-----------
   Total   |     19773          28087  |      47860

Classified + if predicted Pr(D) >= .5
True D defined as _Iworks_1 != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   63.56%
Specificity                     Pr( -|~D)   74.33%
Positive predictive value       Pr( D| +)   63.54%
Negative predictive value       Pr(~D| -)   74.34%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   25.67%
False - rate for true D         Pr( -| D)   36.44%
False + rate for classified +   Pr(~D| +)   36.46%
```

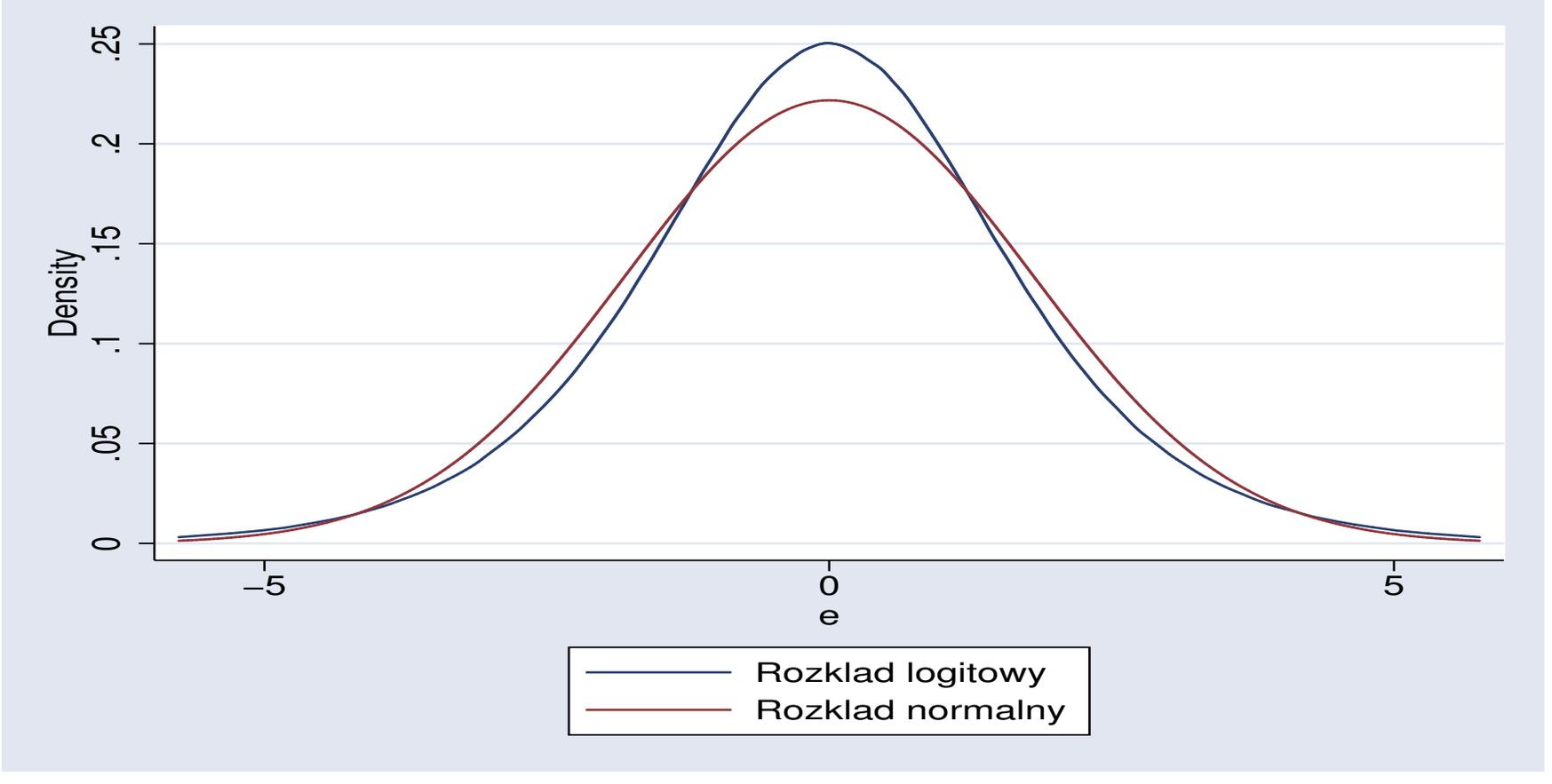
```
False - rate for classified -   Pr( D| -)   25.66%
------------------------------------------------
Correctly classified                          69.88%
------------------------------------------------
```

- To na którą z tych policzonych częstości powinniśmy patrzeć zależy od zastosowania:
  - czy szczególnie ważne jest wyeliminowanie potencjalnych porażek (np. kontrola jakości)
  - czy szczególnie ważne jest uniknięcie fałszywej klasyfikacji (np. identyfikacja przestępców)

Różnice między probitem a logitem

- Inna interpretacja ale intepretacja efektow cząstkowych taka sama

- Oba rozkłady prawdopodobieństwa są symetryczne jednak rozkład logistyczny ma nieco grubsze ogony.



- W związku z tym istotne różnice między modelami będą powstawać dla prób, o nikłym odsetku odpowiedzi 0 albo odpowiedzi 1 i bardzo

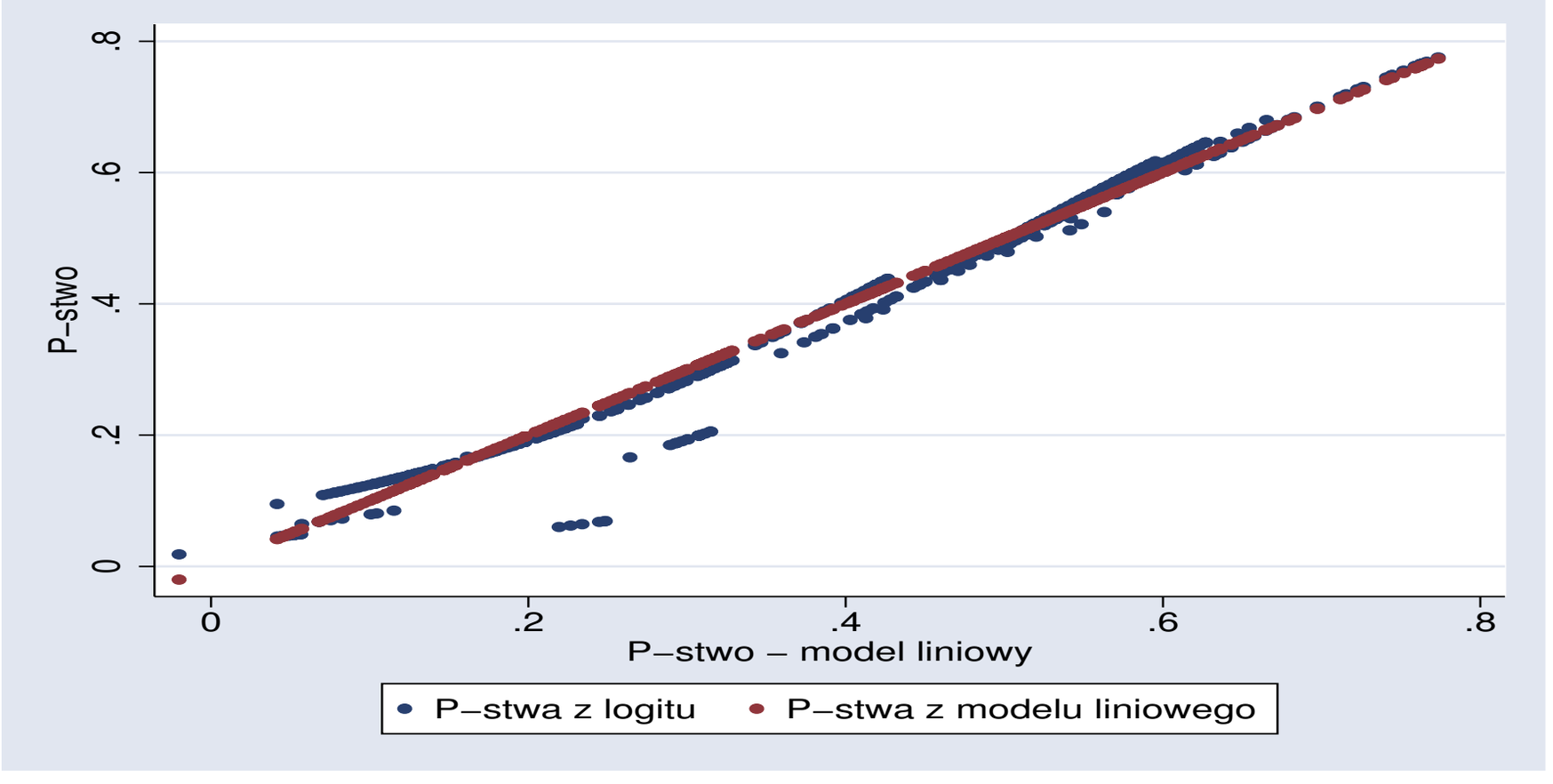
zróżnicowanych zmiennych niezależnych.

- Dla $\overline{x}\beta$ bliskiego 0 funkcja gęstości $f(0) \approx .4$ dla logitu, $f(0) \approx .25$ dla probitu, i $g(0) = 1$ dla LPM. Z tego powodu często zdarza się, że $\beta_{\text{logit}} \approx 1.6\beta_{\text{probit}}$ i $\beta_{\text{logit}} \approx 4\beta_{\text{LPM}}$.

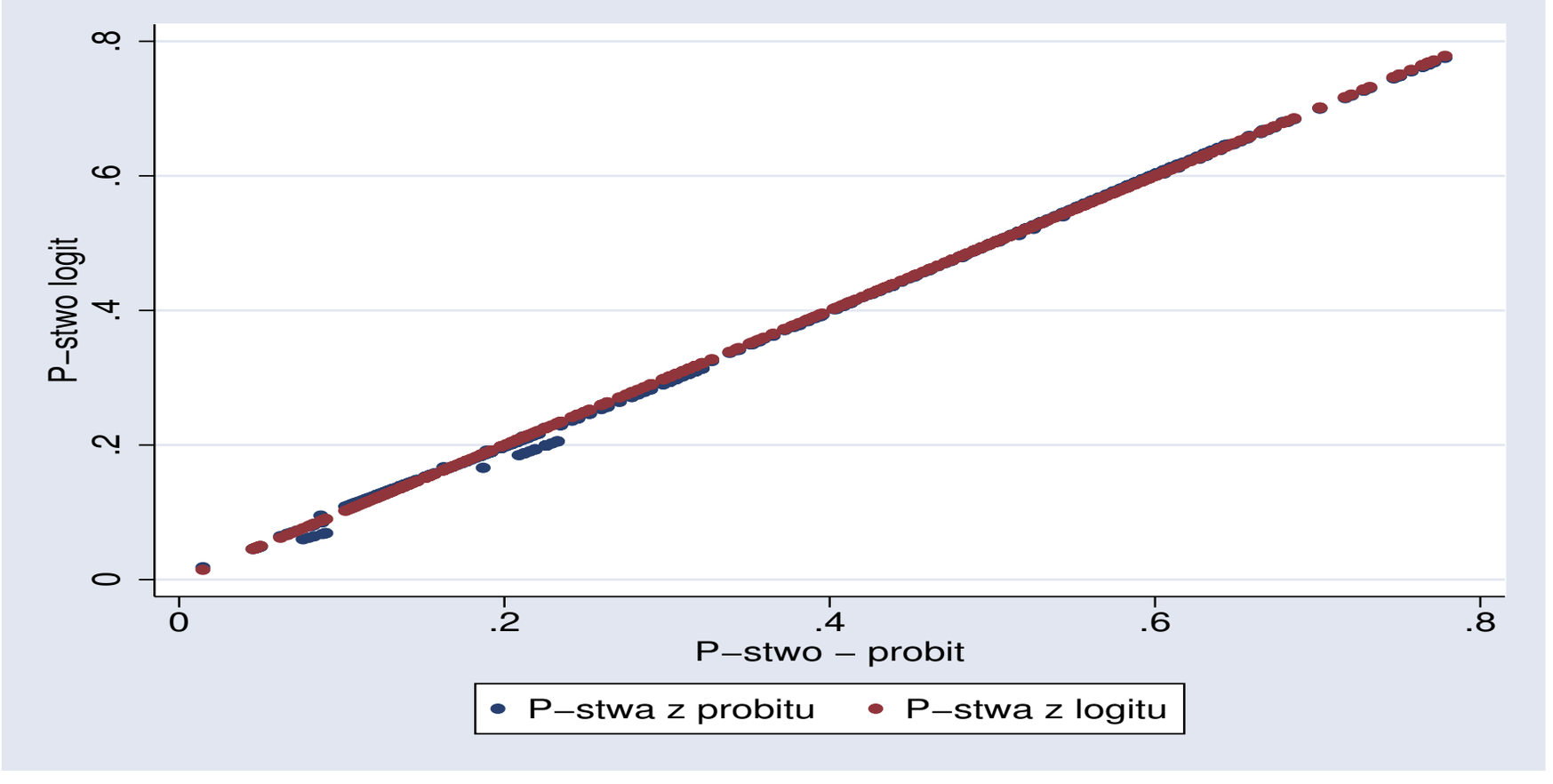
- Nie ma dobrej statystyki, która magłaby posłużyć do wyboru między tymi modelami.

- W praktyce wybieramy ten, który jest analitycznie bardziej wygodny.

- Porównanie przewidywanych prawdopodobieństw z modelu liniowego i logitowego



- Porównanie przewidywanych prawdopodobieństw z modelu probitowego i logitowego



Modele dla zmiennych uporządkowanych

- Zmienna zależna przyjmuje wartości ze zbioru $\{0, 1, 2, \ldots, J\}$ i J jest z góry znane

- Wartości zmiennej zależnej kodują stany, które można w sposób naturalny uporządkować (n.p. oceny z egzaminu, poziomy wykształcenia itp.).

- Uporządkowany probit

$$y^* = x\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

  i dystrybuanta ε jest dana przez $F(\varepsilon)$. Zakładamy, że istnieje J

nieznanych punktów przecięć $\boldsymbol{\alpha}_1 < \boldsymbol{\alpha}_2 < \ldots < \boldsymbol{\alpha}_j$

$$\begin{array}{lll} y = 0 & \text{jeśli} & y^* \leq \boldsymbol{\alpha}_1 \\ y = 1 & \text{jeśli} & \boldsymbol{\alpha}_1 < y^* \leq \boldsymbol{\alpha}_2 \\ \vdots & & \vdots \\ y = J & \text{jeśli} & y^* > \boldsymbol{\alpha}_J \end{array}$$

Prawdopodobieństwa poszczególnych wyborów są równe

$$\begin{aligned} \Pr(y = 0|\,\boldsymbol{x}) &= \Pr(y^* \leq \boldsymbol{\alpha}_1|\,\boldsymbol{x}) = F(\boldsymbol{\alpha}_1 - \boldsymbol{x}\boldsymbol{\beta}) \\ \Pr(y = 1|\,\boldsymbol{x}) &= \Pr(\boldsymbol{\alpha}_1 < y^* \leq \boldsymbol{\alpha}_2|\,\boldsymbol{x}) = F(\boldsymbol{\alpha}_2 - \boldsymbol{x}\boldsymbol{\beta}) - F(\boldsymbol{\alpha}_1 - \boldsymbol{x}\boldsymbol{\beta}) \\ &\vdots \\ \Pr(y = J|\,\boldsymbol{x}) &= \Pr(y^* > \boldsymbol{\alpha}_1|\,\boldsymbol{x}) = 1 - F(\boldsymbol{\alpha}_J - \boldsymbol{x}\boldsymbol{\beta}) \end{aligned}$$

- Te prawdopobieństwa możemy użyć do zdefiniowania funkcji wiarygodności

- Dla uporządkowanego probitu $F(\varepsilon) = \Phi(\varepsilon)$ a dla uporządkowanego logitu $F(\varepsilon) = \Lambda(\varepsilon)$

- Dla modelu dla zmiennych uporządkowanych efekt cząstkowy jest równy

$$\begin{aligned}
\frac{\partial p_0(\boldsymbol{x})}{\partial \boldsymbol{x}} &= -\boldsymbol{\beta}_k f(\boldsymbol{\alpha}_1 - \boldsymbol{x}\boldsymbol{\beta}), \ \frac{\partial p_J(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{\beta}_k f(\boldsymbol{\alpha}_J - \boldsymbol{x}\boldsymbol{\beta}) \\
\frac{\partial p_j(\boldsymbol{x})}{\partial \boldsymbol{x}} &= \boldsymbol{\beta}_k \left[f(\boldsymbol{\alpha}_{j-1} - \boldsymbol{x}\boldsymbol{\beta}) - f(\boldsymbol{\alpha}_j - \boldsymbol{x}\boldsymbol{\beta}) \right], \quad 0 < j < J
\end{aligned}$$

- Znaki efektów cząskowych są zdeterminowane przez znaki $\boldsymbol{\beta}$ tylko dla $j = 1$ i $j = J$. Na przykład jeśli $\boldsymbol{\beta}_k$ jest dodatnie to wpływ x_k na prawdopodobieństwo alternatywy 0 jest ujemny a na prawdodpodobieństwo alternatywy J dodatni.

- Dla pośrednich wyborów znak zależy także od znaku różnicy $\boldsymbol{\phi}(\boldsymbol{\alpha}_{j-1} - \boldsymbol{x}\boldsymbol{\beta}) - \boldsymbol{\phi}(\boldsymbol{\alpha}_j - \boldsymbol{x}\boldsymbol{\beta})$

- Dla tych modeli możemy także policzyć procenty prawidłowych przewidywań (przewidywany wybór byłby tu wyborem najabrdziej prawdopodobnym według modelu)

Przykład Zależność wykształcenia od wieku, wielkości miejscowości (w której respondent mieszkał mając 14 lat) i wykształcenia ojca - dane PGSS. Zmienną zelażną jest poziom wykształcenia kodowany od 0 do 9 zgodnie rosnąco wraz z poziomem wykształcenia. Tak samo rosnąco zakodowany jest wielkość miejscowości. Badane są jedynie osoby w wiku powyżej 25 lat.

- Wielkości parametrów:

```
Ordered probit estimates                          Number of obs   =       6842
                                                  LR chi2(17)     =    2957.65
                                                  Prob > chi2     =     0.0000
Log likelihood = -11395.434                       Pseudo R2       =     0.1149

------------------------------------------------------------------------------
      degree |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   -.023434   .0009754   -24.02   0.000    -.0253458   -.0215222
   _Ires14_2 |   .4703715   .0513952     9.15   0.000     .3696387    .5711043
   _Ires14_3 |   .4559041   .0506483     9.00   0.000     .3566353    .5551729
   _Ires14_4 |   .4556405   .0575138     7.92   0.000     .3429154    .5683655
   _Ires14_5 |   .4827543   .0622395     7.76   0.000     .3607672    .6047414
   _Ires14_6 |   .4322438   .0560347     7.71   0.000     .3224177    .5420698
   _Ires14_7 |   .4563276   .0663018     6.88   0.000     .3263785    .5862766
   _Ires14_8 |   .5490264   .0566223     9.70   0.000     .4380488    .6600039
   _Ipadeg_1 |   .5281177   .0748967     7.05   0.000     .3813229    .6749124
   _Ipadeg_2 |   .9844236   .0748935    13.14   0.000      .837635    1.131212
   _Ipadeg_3 |   1.130091   .0822737    13.74   0.000     .9688372    1.291344
   _Ipadeg_4 |   1.823293   .1148838    15.87   0.000     1.598125    2.048461
```

```
   _Ipadeg_5 |   1.740986   .1067116    16.31   0.000     1.531835    1.950136
   _Ipadeg_6 |   1.640127   .0907487    18.07   0.000     1.462263    1.817991
   _Ipadeg_7 |   2.138463    .190953    11.20   0.000     1.764202    2.512724
   _Ipadeg_8 |   2.124689   .1958153    10.85   0.000     1.740898     2.50848
   _Ipadeg_9 |     2.3626   .1120992    21.08   0.000     2.142889     2.58231
-------------+----------------------------------------------------------------
        _cut1 |  -3.273454    .113341          (Ancillary parameters)
        _cut2 |  -1.932496   .0969967
        _cut3 |  -.4900235   .0935463
        _cut4 |   .3189546    .093053
        _cut5 |   .3916605   .0931025
        _cut6 |   .5746352    .093201
        _cut7 |   1.340311   .0940497
        _cut8 |    1.58867   .0946985
        _cut9 |   1.691446    .095045
------------------------------------------------------------------------------
```

- Wnioski:
  - wszystkie analizowane zmienne mają istotny wpływ na zmienną zależną
  - ujemny wpływ na p-stwo posiadania najniższego poziomu wykształcenia

i dodatni na p-stwo uzyskania najwyższego wykształcenia ma zmienna res14 (wilkość miejscowości), przy czym wydaje się, że istotne znaczenia ma jedynie to, czy respondent pochodzi ze wsi, ze zwykłego miasta, czy z dużego miasta

– ujemny wpływ na p-stwo posiadania najniższego poziomu wykształcenia i i dodatni na p-stwo uzyskania najwyższego wykształcenia ma zmienna padeg - poziom wykształcenia wydaje się do pewnego stopnia dziedziczny

- Wielkości efektów cząstkowych dla poziomu wykształcenia 2 (podstawowe)

```
Marginal effects after oprobit
      y  = Pr(degree==2) (predict, outcome(2))
         =  .26783642
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |   .0067511      .00031   21.65   0.000    .00614  .007362   48.3965
_Ires1~2*|  -.1233111      .01193  -10.34   0.000   -.14669 -.099932   .069424
_Ires1~3*|  -.1201262      .01189  -10.11   0.000  -.143424 -.096828   .074247
_Ires1~4*|  -.1195408      .01332   -8.98   0.000  -.145639 -.093442    .05627
_Ires1~5*|  -.1253851      .01402   -8.95   0.000  -.152857 -.097913   .047208
_Ires1~6*|  -.1142518      .01323   -8.64   0.000  -.140174  -.08833   .060801
_Ires1~7*|  -.1192819      .01518   -7.86   0.000  -.149026 -.089538   .042239
_Ires1~8*|  -.1404233      .01233  -11.39   0.000  -.164596 -.116251   .061532
_Ipade~1*|  -.1422198      .01857   -7.66   0.000  -.178608 -.105832   .228588
_Ipade~2*|  -.2640863      .01846  -14.31   0.000   -.30026 -.227913   .440076
_Ipade~3*|  -.2520474      .01296  -19.44   0.000  -.277458 -.226637   .141771
_Ipade~4*|  -.2693371      .00666  -40.46   0.000  -.282383 -.256291   .021193
_Ipade~5*|  -.2695223      .00692  -38.93   0.000  -.283092 -.255953    .02777
_Ipade~6*|  -.2802689      .00782  -35.84   0.000  -.295595 -.264943   .064309
_Ipade~7*|  -.2676569      .00617  -43.35   0.000  -.279758 -.255556     .0057
```

```
_Ipad~_8*|  -.2673083      .00621  -43.01   0.000  -.279488 -.255128   .005408
_Ipade~9*|  -.2859779      .00617  -46.35   0.000  -.298071 -.273884    .02967
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

- Intepretacja parametrów:
  - dla _Ipade~9 - respondent, którego ojciec ma wykształcenie wyższe ma o 29% niższe prawdopodbieństwo uzyskania wykształcenia jedynie podstawowego niż osoba, której ojciec nie miał żadnego wykształcenia (poziom 0 - bazowy)
  - wraz ze wzrostem wieku o 1, p-stwo uzyskania wykształcenia podstawowego rośnie o 0.6%

Modele dla liczebności

- Liczebność jest to zmienna, która przyjmuje nieujemne wartości calkowite (np. liczba dzieci, liczba strajków ciągu roku w danej firmie, liczba papierosów wypalanych dziennie itd.)

- Czasami liczebność może mieć jakąś (zależną od obserwacji) logiczną górną granice (np. liczba dzieci danej w rodzinie, które ukończyły wyższe studia nie może być wyższa niż ogólna liczba dzieci w tej rodzinie)

- W tych przypadkach zastosowanie modelu dla zmiennych uporządkowanych jest nieuprawnione, ponieważ zbiór możliwych wartości y nie jest z góry określony (naogół wyniki estymacji uzyskane tą metodą nie będą przy tym satysfakcjonujące)

- Najprostsza metoda estymacji w przypadku zmiennych to założyć, że $\mathrm{E}\left(y|\,\boldsymbol{x}\right)=\boldsymbol{x\beta}$ i estymować model przy użyciu OLS.

- Wadą tej metody jest fakt, że niektóre z dopasowanych $\widehat{y}=\boldsymbol{xb}$ mogą się okazać ujemne, co nie ma sensu, ponieważ liczebnosć musi być zawsze dodatnia.

- Inna możliwość to dokonanie na zmiennych przekształcenia logarytmicznego i założenie, że $\log\mathrm{E}\left[y|\,\boldsymbol{x}\right]=\boldsymbol{x\beta}$. Wartości dopasowane są w tym modelu równe $\widehat{y}=\exp\left(\boldsymbol{xb}\right)>0$ i większe od zera. Te podejście jest jednak niemożliwe do zastosowanie jeśli dla istotnej części obserwacji $y=0$ (np. liczba dzieci)

Model Poissona

- Najpopularniejszym modelem dla liczebności jest model Poissona
- Zakładamy, że warunkowa wartość oczekiwana y jest dana przez $\mathrm{E}\left(y|\,\boldsymbol{x}\right)=\boldsymbol{\mu}\left(\boldsymbol{x}\right)$ (najczęściej $\boldsymbol{\mu}\left(\boldsymbol{x}\right)=\exp\left(\boldsymbol{x\beta}\right)$)
- Warunkowa dystrybuanta y jest dane przez rozkład Poissona dla $\boldsymbol{\lambda}=\boldsymbol{\mu}\left(\boldsymbol{x}\right)$

$$f\left(y|\,\boldsymbol{x}\right)=\frac{e^{-\boldsymbol{\lambda}}\boldsymbol{\lambda}^{y}}{y!}=\frac{\exp\left[-\boldsymbol{\mu}\left(\boldsymbol{x}\right)\right]\left[\boldsymbol{\mu}\left(\boldsymbol{x}\right)\right]^{y}}{y^{!}}$$

- Rzeczywiście dla rozkładu Poissona $\mathrm{E}\left(y|\,\boldsymbol{x}\right)=\boldsymbol{\lambda}=\boldsymbol{\mu}\left(\boldsymbol{x}\right)$

- Efekty cząstkowe w rozkładzie Poissona z $\boldsymbol{\mu}(\boldsymbol{x}) = \exp(\boldsymbol{x}\boldsymbol{\beta})$ intepretujemy tak samo jak w modelu, w którym zmienna zależna jest jest zlogarytmowana a niezależne nie są zlogarytmowane (semielastyczności wartości oczekiwanej)

$$\frac{\partial \ln \mathrm{E}(y|\,\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial \ln \boldsymbol{\mu}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{\beta}$$

  a więc $\frac{\Delta \boldsymbol{\mu}(\boldsymbol{x})}{\boldsymbol{\mu}(\boldsymbol{x})} \Big/ \Delta x_j = \boldsymbol{\beta}_j$.

- Jeśli zmienne niezależne są także zlogarytmowane to parametry będziemy intepretować jako elastyczności wartości oczekiwanej.

- Parametry tego modelu mogą być zgodnie i efektywnie oszacowane za pomocą MNW

- **Najważniejsze ograniczenie modelu Poissona**. Dla dystrybuanty

rozkładu Poissona:

$$\mathrm{E}\left(y|\,\boldsymbol{x}\right)=\mathrm{Var}\left(y|\,\boldsymbol{x}\right)=\boldsymbol{\lambda}=\boldsymbol{\mu}\left(\boldsymbol{x}\right)$$

- Założenie to bardzo często nie jest spełnione w praktyce (rzadko kiedy można na podstawie rozważań teoretycznych uzasadnić rowność wartości oczekiwanej i wariancji)

- Założenie o równości wartości oczekiwanej i wariancji można przetestować.

- Można pokazać, jeśli tylko dobrze jest wyspecyfikowany model dla warunkowej wartości oczekiwanej ($\mathrm{E}\left(y|\,\boldsymbol{x}\right)=\boldsymbol{\mu}\left(\boldsymbol{x}\right)$), to nawet dla $\mathrm{Var}\left(y|\,\boldsymbol{x}\right)\neq\boldsymbol{\mu}\left(\boldsymbol{x}\right)$, model Poissona będzie dawał zgodne oszacowania parametrów.

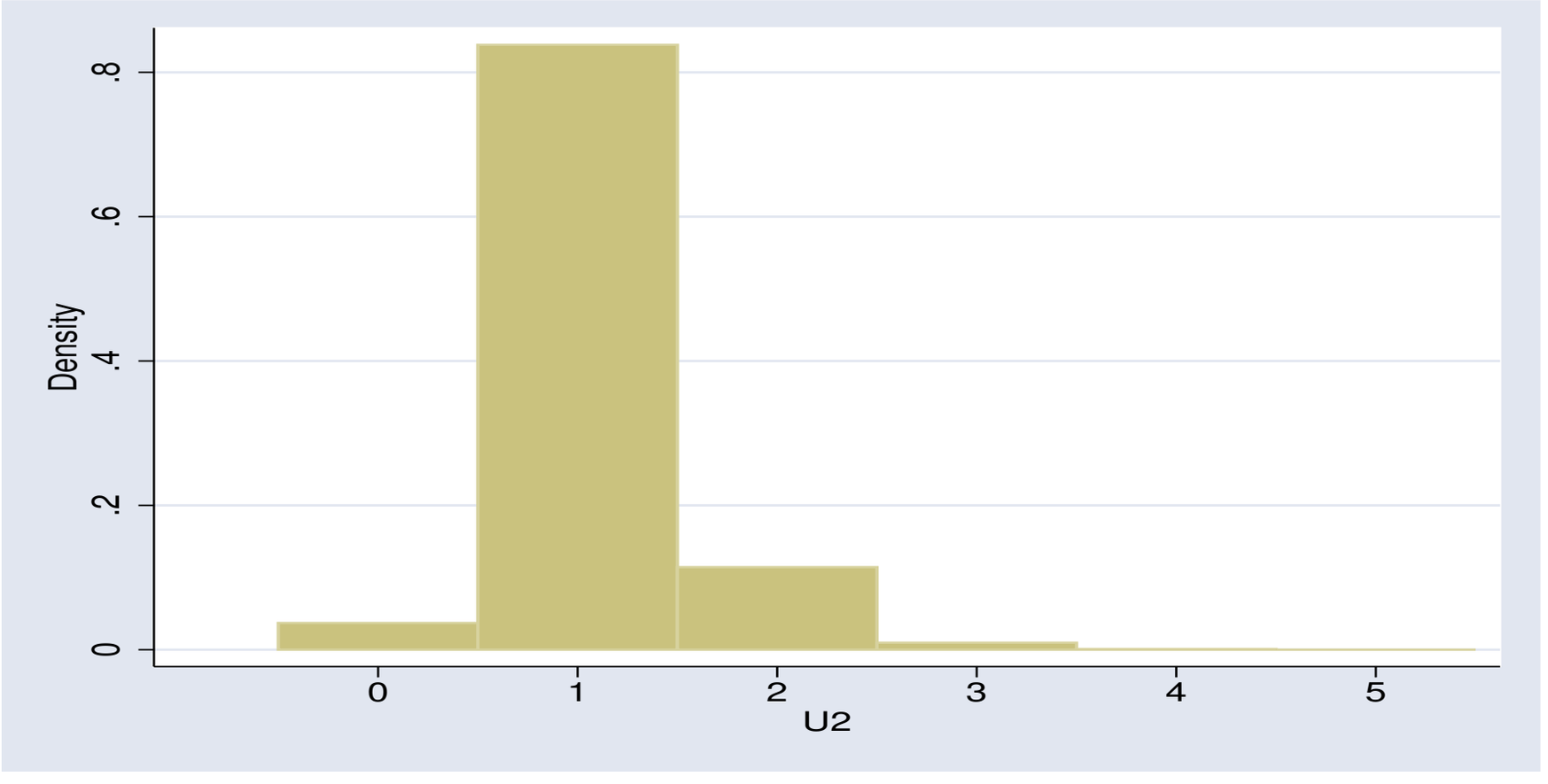
- W takim przypadku oszacowanie macierzy wariancji kowariancji będzie

jednak nieprawidłowe. Można jednak w tym przypadku zastosować macierz odporną wariancji kowariancji ($Pseudo - MNW$)

- Czasami zdarza się, że obserwujemy zbyt wysoką liczbę zer w danych (np. przy liczbie wypalanych papierosów wiąże się ro z tym, że ludzie napierw wybierają palenie bądź nie a później, to ile będą palić). W takim przypadku powinno się stosować model ZIP (**Z**ero **I**nflated **P**oisson - rozdęty w zerze model Poissona)

Przykład Liczba telewizorów w gospodarstwie domowym a logarytm dochodu i liczba osób w rodzinie. Założono, że $\boldsymbol{\mu}(\boldsymbol{x}) = \exp(\boldsymbol{x}\boldsymbol{\beta})$.

- Liczba telewizorów w gospodarstwie domowym (dane z budżetów gospodarstw domowych)


Density
0
.2
.4
.6
.8
0
1
2
3
4
5
U2

```
Poisson regression                                Number of obs   =      35952
                                                  LR chi2(7)      =     655.67
                                                  Prob > chi2     =     0.0000
Log likelihood = -39036.892                       Pseudo R2       =     0.0083

------------------------------------------------------------------------------
          tv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       ldochg |     .15787     .00923    17.10   0.000     .1397795    .1759604
         los |   .0289983   .0035735     8.11   0.000     .0219944    .0360021
      _Iklm_2 |  -.0091403   .0206934    -0.44   0.659    -.0496986    .0314181
      _Iklm_3 |  -.0225221     .02266    -0.99   0.320    -.0669348    .0218906
      _Iklm_4 |    .001201   .0183397     0.07   0.948    -.0347441    .0371461
      _Iklm_5 |   .0097115   .0199079     0.49   0.626    -.0293072    .0487301
      _Iklm_6 |  -.0873394   .0178009    -4.91   0.000    -.1222285   -.0524503
        _cons |  -1.137438   .0686759   -16.56   0.000    -1.272041   -1.002836
------------------------------------------------------------------------------
```

- Sprawdzamy, czy liczebności pasują do rozkładu Poissona

```
Goodness-of-fit chi2  =  6183.881
Prob > chi2(36155)    =    1.0000
```

- Dobre dopasowanie rozkładu

- Intepretacja parametrów:
  - zwiększenie się dochodu o 1% powoduje zwiększenie wartości oczekiwanej liczby telewizorów w gospodarstwie o 0.15%
  - zwiększenie się liczby osób o 1 zwiększa liczbę telewizorów o 2.9%
  - zamieszkanie na wsi zmniejsza oczekiwaną liczbę telewizorów o 8.7%

- Liczba telewizorów w gospodarstwie domowym (liczebności rzeczywiste i dopasowane)