Binarne zmienne zależne

- Zmienna binarna to zmienna przyjmująca dwie wartości

- Zmienna binarna to zmienna przyjmująca dwie wartości
- Zazwyczaj przyjmuje się, że 1 oznacza "sukces", a 0 "porażkę"

- Zmienna binarna to zmienna przyjmująca dwie wartości
- Zazwyczaj przyjmuje się, że 1 oznacza "sukces", a 0 "porażkę"
- Celem budowy modelu jest oszacowanie warunkowego prawdopodobieństwa wystąpienia jednego ze stanów

- Dla zmiennej binarnej prawdopodobieństwo jest opisane przez model dwumianowy

$$Pr(y_i|X = x_i) = \begin{cases} 1 - Pr(X = x_i) & y_i = 0 \\ Pr(X = x_i) & y_i = 1 \end{cases}$$

- Dla zmiennej binarnej prawdopodobieństwo jest opisane przez model dwumianowy

$$Pr(y_i|X = x_i) = \begin{cases} 1 - Pr(X = x_i) & y_i = 0 \\ Pr(X = x_i) & y_i = 1 \end{cases}$$

- Dla uproszczenia zapisu przyjmijmy, że $X \equiv (X = x_i)$

- Dla zmiennej binarnej prawdopodobieństwo jest opisane przez model dwumianowy

$$Pr(y_i|X = x_i) = \begin{cases} 1 - Pr(X = x_i) & y_i = 0 \\ Pr(X = x_i) & y_i = 1 \end{cases}$$

- Dla uproszczenia zapisu przyjmijmy, że $X \equiv (X = x_i)$
- Więc

$$Pr(y|X) = [P(X)]^{y_i}[1 - P(X)]^{1-y_i}$$

- Forma funkcyjna modelu i postać funkcji wiarogodności zależy od wyboru funkcji $P(X)$

- Forma funkcyjna modelu i postać funkcji wiarogodności zależy od wyboru funkcji $P(X)$
- W zastosowaniach przyjmuje się, że $P(X)$ powinna być ciągła lub jednostronnie ciągła

- Forma funkcyjna modelu i postać funkcji wiarogodności zależy od wyboru funkcji $P(X)$
- W zastosowaniach przyjmuje się, że $P(X)$ powinna być ciągła lub jednostronnie ciągła
- Zazwyczaj przyjmuje się, że $P(X) = F(X_i\beta)$, gdzie $F(\cdot)$ jest dystrybuantą

- Forma funkcyjna modelu i postać funkcji wiarogodności zależy od wyboru funkcji $P(X)$
- W zastosowaniach przyjmuje się, że $P(X)$ powinna być ciągła lub jednostronnie ciągła
- Zazwyczaj przyjmuje się, że $P(X) = F(X_i\beta)$, gdzie $F(\cdot)$ jest dystrybuantą
- Dodatkowo często zakłada się że model jest liniowy względem argumentów dystrybuanty

- Funkcja wiarogodnośći modelu dla binanrnej zmiennej zależnej ma postać

$$L(\theta) = \prod_{i=1}^{N} [P(X)]^{y_i} [1 - P(X)]^{1-y_i}$$

- Funkcja wiarogodnośći modelu dla binanrnej zmiennej zależnej ma postać

$$L(\theta) = \prod_{i=1}^{N} [P(X)]^{y_i} [1 - P(X)]^{1-y_i}$$

- Zatem jej logarytm

$$\ell(\theta) = \sum_{i=1}^{N} y_i \ln P(X_i) + (1 - y_i) \ln(1 - P(X))$$

- Funkcja wiarogodnośći modelu dla binanrnej zmiennej zależnej ma postać

$$L(\theta) = \prod_{i=1}^{N} [P(X)]^{y_i} [1 - P(X)]^{1-y_i}$$

- Zatem jej logarytm

$$\ell(\theta) = \sum_{i=1}^{N} y_i \ln P(X_i) + (1 - y_i) \ln(1 - P(X))$$

- ponieważ $y_i \in \{0, 1\}$, $P(X_i) \in [0, 1]$, więc $\ln P(X_i) \leq 0$ oraz $\ln(1 - P(X_i)) \leq 0$, zatem $\ell(\theta) \leq 0$

- Zakładamy istnienie nieobserwowanej zmiennej (indeksu)

- Zakładamy istnienie nieobserwowanej zmiennej (indeksu)
- Determinuje ona wartość zmiennej zależnej

- Zakładamy istnienie nieobserwowanej zmiennej (indeksu)
- Determinuje ona wartość zmiennej zależnej
- Przyjmuje się, że zmienna zależna jest realizacją procesu

$$y_i^\star = X_i\beta + \varepsilon_i \qquad E(\varepsilon_i) = 0 \quad \varepsilon_i \text{ jest symetryczny}$$

- Zakładamy istnienie nieobserwowanej zmiennej (indeksu)
- Determinuje ona wartość zmiennej zależnej
- Przyjmuje się, że zmienna zależna jest realizacją procesu

$$y_i^\star = X_i\beta + \varepsilon_i \qquad E(\varepsilon_i) = 0 \quad \varepsilon_i \text{ jest symetryczny}$$

- Obserwowana wartość zależy od wartości zmiennej ukrytej

$$y_i = \begin{cases} 0 \text{ gdy} & y_i^\star < 0 \\ 1 \text{ gdy} & y_i^\star \geq 0 \end{cases}$$

- Zatem prawdopodobieństwo sukcesu wynosi

$$Pr(y_i = 1) = Pr(y_i^\star \geq 0) = Pr(X_i\beta + \varepsilon_i \geq 0) = Pr(\varepsilon_i > -X_i\beta)$$

- Zatem prawdopodobieństwo sukcesu wynosi

$$Pr(y_i = 1) = Pr(y_i^\star \geq 0) = Pr(X_i\beta + \varepsilon_i \geq 0) = Pr(\varepsilon_i > -X_i\beta)$$

- Z symetryczności rozkładu ε uzyskujemy

$$Pr(\varepsilon_i > -X_i\beta) = Pr(\varepsilon_i < X_i\beta) = F(X_i\beta)$$

- Podejście indeksowe stosowane jest, gdy wartości zmiennej zależnej modelu nie mogą być interpretowane jako decyzje jednostek, a zmienne objaśniające to charakterystyki jednostek

- Podejście indeksowe stosowane jest, gdy wartości zmiennej zależnej modelu nie mogą być interpretowane jako decyzje jednostek, a zmienne objaśniające to charakterystyki jednostek
- Indeks nie posiada interpretacji ekonomicznej

- Zakładamy, że jednostki są racjonalne i zachowują się zgodnie z teorią użyteczności

- Zakładamy, że jednostki są racjonalne i zachowują się zgodnie z teorią użyteczności
- Jednostka i wybiera jedną spośród J dostępnych alternatyw

- Zakładamy, że jednostki są racjonalne i zachowują się zgodnie z teorią użyteczności
- Jednostka i wybiera jedną spośród J dostępnych alternatyw
- Z każdą alternatywą związany jest pewien poziom użyteczności

- Zakładamy, że jednostki są racjonalne i zachowują się zgodnie z teorią użyteczności
- Jednostka i wybiera jedną spośród J dostępnych alternatyw
- Z każdą alternatywą związany jest pewien poziom użyteczności
- Użyteczność jednostki i z wyboru alternatywy j wynosi

$$U_{ij} = X_{ij} + \varepsilon_{ij}$$

- Zakładamy, że jednostki są racjonalne i zachowują się zgodnie z teorią użyteczności
- Jednostka i wybiera jedną spośród J dostępnych alternatyw
- Z każdą alternatywą związany jest pewien poziom użyteczności
- Użyteczność jednostki i z wyboru alternatywy j wynosi

$$U_{ij} = X_{ij} + \varepsilon_{ij}$$

- Użyteczność zależy od ujawnionych charakterystyk X_{ij}, oraz cech nieznanych badaczowi ε_{ij}

- Zakłada się, że część ujawniona jest wszystkim znana. Składa się z charakterystyk decydentów i/lub alternatyw

- Zakłada się, że część ujawniona jest wszystkim znana. Składa się z charakterystyk decydentów i/lub alternatyw
- Cześć nieujawniona jest znana tylko jednostce podejmującej decyzję

- Zakłada się, że część ujawniona jest wszystkim znana. Składa się z charakterystyk decydentów i/lub alternatyw
- Cześć nieujawniona jest znana tylko jednostce podejmującej decyzję
- Dodatkowo zakłada się, że obie części są warunkowo niezależne

- Zakłada się, że część ujawniona jest wszystkim znana. Składa się z charakterystyk decydentów i/lub alternatyw
- Cześć nieujawniona jest znana tylko jednostce podejmującej decyzję
- Dodatkowo zakłada się, że obie części są warunkowo niezależne
- Decydent maksymalizuje swoją użyteczność następująco

$$Pr(y = j) = Pr(U_j > U_k \; \forall j \neq k) = Pr(X_j\beta_j + \varepsilon_j > X_k\beta_k + \varepsilon_k) =$$

- Zakłada się, że część ujawniona jest wszystkim znana. Składa się z charakterystyk decydentów i/lub alternatyw
- Cześć nieujawniona jest znana tylko jednostce podejmującej decyzję
- Dodatkowo zakłada się, że obie części są warunkowo niezależne
- Decydent maksymalizuje swoją użyteczność następująco

$$Pr(y = j) = Pr(U_j > U_k \; \forall j \neq k) = Pr(X_j\beta_j + \varepsilon_j > X_k\beta_k + \varepsilon_k) =$$

$$= Pr(\varepsilon_k - \varepsilon_j < X_j\beta_j - X_k\beta_k) = F(X_j\beta_j - X_k\beta_k)$$

- Zakłada się, że część ujawniona jest wszystkim znana. Składa się z charakterystyk decydentów i/lub alternatyw
- Cześć nieujawniona jest znana tylko jednostce podejmującej decyzję
- Dodatkowo zakłada się, że obie części są warunkowo niezależne
- Decydent maksymalizuje swoją użyteczność następująco

$$Pr(y = j) = Pr(U_j > U_k \; \forall j \neq k) = Pr(X_j\beta_j + \varepsilon_j > X_k\beta_k + \varepsilon_k) =$$

$$= Pr(\varepsilon_k - \varepsilon_j < X_j\beta_j - X_k\beta_k) = F(X_j\beta_j - X_k\beta_k)$$

- Analityczna postać modelu zależy od rozkładu $F(\cdot)$, charakterystyk alternatywy wybranej j oraz alternatyw niewybranych k

- Modele dla zmiennych binarnych i dyskretnych nie są liniowe, ponieważ

$$E(y|X) = 0[1 - F(X\beta)] + 1F(X\beta) = F(X\beta) = Pr(X\beta)$$

- Modele dla zmiennych binarnych i dyskretnych nie są liniowe, ponieważ

$$E(y|X) = 0[1 - F(X\beta)] + 1F(X\beta) = F(X\beta) = Pr(X\beta)$$

- Wartość oczekiwana zmiennej zależnej jest równa prawdopodobieństwu sukcesu

- Modele dla zmiennych binarnych i dyskretnych nie są liniowe, ponieważ

$$E(y|X) = 0[1 - F(X\beta)] + 1F(X\beta) = F(X\beta) = Pr(X\beta)$$

- Wartość oczekiwana zmiennej zależnej jest równa prawdopodobieństwu sukcesu
- Model będzie liniowy wtedy i tylko wtedy, gdy prawdopodobieństwo jest funkcją liniową zmiennych i parametrów

- Modele dla zmiennych binarnych i dyskretnych nie są liniowe, ponieważ

$$E(y|X) = 0[1 - F(X\beta)] + 1F(X\beta) = F(X\beta) = Pr(X\beta)$$

- Wartość oczekiwana zmiennej zależnej jest równa prawdopodobieństwu sukcesu
- Model będzie liniowy wtedy i tylko wtedy, gdy prawdopodobieństwo jest funkcją liniową zmiennych i parametrów
- Z uwagi na nieliniowość modelu interpretuje się efekty cząstkowe

- Efekt cząstkowy to pochodna funkcji wartości oczekiwanej zmiennej zależnej względem wartości zmiennych objasniających

$$\frac{\partial E(y|X)}{\partial x_k} = \frac{\partial Pr(y=1|X)}{\partial x_k} = f(X\beta)\beta_k$$

- gdzie $f(\cdot)$ jest funkcją łącznej gęstości

- Efekt cząstkowy to pochodna funkcji wartości oczekiwanej zmiennej zależnej względem wartości zmiennych objasniających

$$\frac{\partial E(y|X)}{\partial x_k} = \frac{\partial Pr(y=1|X)}{\partial x_k} = f(X\beta)\beta_k$$

- gdzie $f(\cdot)$ jest funkcją łącznej gęstości
- Efekt cząstkowy interpretuje się jako wpływ jednostkowej zmiany zmiennej niezależnej na wielkość prawdopodobieństwa sukcesu

- Efekt cząstkowy to pochodna funkcji wartości oczekiwanej zmiennej zależnej względem wartości zmiennych objasniających

$$\frac{\partial E(y|X)}{\partial x_k} = \frac{\partial Pr(y=1|X)}{\partial x_k} = f(X\beta)\beta_k$$

- gdzie $f(\cdot)$ jest funkcją łącznej gęstości
- Efekt cząstkowy interpretuje się jako wpływ jednostkowej zmiany zmiennej niezależnej na wielkość prawdopodobieństwa sukcesu
- Jego wartość zależy od wartości wektora x

- Efekt cząstkowy to pochodna funkcji wartości oczekiwanej zmiennej zależnej względem wartości zmiennych objasniających

$$\frac{\partial E(y|X)}{\partial x_k} = \frac{\partial Pr(y=1|X)}{\partial x_k} = f(X\beta)\beta_k$$

- gdzie $f(\cdot)$ jest funkcją łącznej gęstości
- Efekt cząstkowy interpretuje się jako wpływ jednostkowej zmiany zmiennej niezależnej na wielkość prawdopodobieństwa sukcesu
- Jego wartość zależy od wartości wektora x
- Z reguły efekty cząstkowe są liczone dla średnich wartości zmiennych egzogenicznych

- W przypadku dyskretnych zmiennych objaśniających pochodna nie istnieje

- W przypadku dyskretnych zmiennych objaśniających pochodna nie istnieje
- Z tego powodu oblicza się tzw. zmiany dyskretne

$$\frac{\Delta E(y|X)}{\Delta x_k} = Pr(y = 1|X, x_k = 1) - Pr(y = 1|X, x_k = 0)$$

- W przypadku dyskretnych zmiennych objaśniających pochodna nie istnieje
- Z tego powodu oblicza się tzw. zmiany dyskretne

$$\frac{\Delta E(y|X)}{\Delta x_k} = Pr(y = 1|X, x_k = 1) - Pr(y = 1|X, x_k = 0)$$

- Efekt cząstkowy interpretuje się jako zmianę prawdopodobieństwa sukcesu pod wpływem zmiany wartości zmiennej niezależnej z 0 na 1

- W przypadku dyskretnych zmiennych objaśniających pochodna nie istnieje
- Z tego powodu oblicza się tzw. zmiany dyskretne

$$\frac{\Delta E(y|X)}{\Delta x_k} = Pr(y = 1|X, x_k = 1) - Pr(y = 1|X, x_k = 0)$$

- Efekt cząstkowy interpretuje się jako zmianę prawdopodobieństwa sukcesu pod wpływem zmiany wartości zmiennej niezależnej z 0 na 1
- Ponieważ wartość funkcji gęstości jest nieujemna to znak efektu cząstkowego jest taki sam jak znak oszacowanej wartości współczynnika modelu

- Jest to najprostszy model

$$P(X_i) = F(X_i\beta) = X_i\beta$$

- Jest to najprostszy model

$$P(X_i) = F(X_i\beta) = X_i\beta$$

- Model jest liniowy ponieważ

$$E(y_i|X_i) = X_i\beta$$

- Jest to najprostszy model

$$P(X_i) = F(X_i\beta) = X_i\beta$$

- Model jest liniowy ponieważ

$$E(y_i|X_i) = X_i\beta$$

- Estymatory MNK dla parametrów modelu są nieobciażone, poniewaz

$$E(\varepsilon_i|X_i) = E(y_i = X_i\beta|X_i) = X_i\beta - X_i\beta = 0$$

- Jest to najprostszy model

$$P(X_i) = F(X_i\beta) = X_i\beta$$

- Model jest liniowy ponieważ

$$E(y_i|X_i) = X_i\beta$$

- Estymatory MNK dla parametrów modelu są nieobciażone, poniewaz

$$E(\varepsilon_i|X_i) = E(y_i = X_i\beta|X_i) = X_i\beta - X_i\beta = 0$$

- W modelu liniowym można ilościowo interpretować parametry, ponieważ

$$\frac{\partial E(y|X)}{\partial x_k} = \frac{\partial X\beta}{\partial x_k} = \beta_k$$

- Jednak LMP posiada liczne wady

- Jednak LMP posiada liczne wady
  1. Nie ma gwarancji, że $\hat{y}_i \in [0, 1]$, a wartości z poza przedziału nie mogą być interpretowane jako prawdopodobieństwa

- Jednak LMP posiada liczne wady
  1. Nie ma gwarancji, że $\hat{y}_i \in [0, 1]$, a wartości z poza przedziału nie mogą być interpretowane jako prawdopodobieństwa
  2. Składnik losowy modelu jest heteroscedastyczny

$$\mathrm{var}(y_i|X) = E(y_i^2|X) = [E(y_i|X)]^2 = X_i\beta - (X_i\beta)^2 = X_i\beta(1 - X_i\beta)$$

- Jednak LMP posiada liczne wady
  1. Nie ma gwarancji, że $\hat{y}_i \in [0, 1]$, a wartości z poza przedziału nie mogą być interpretowane jako prawdopodobieństwa
  2. Składnik losowy modelu jest heteroscedastyczny

$$\text{var}(y_i|X) = E(y_i^2|X) = [E(y_i|X)]^2 = X_i\beta - (X_i\beta)^2 = X_i\beta(1 - X_i\beta)$$

- Zatem wariancja składnika losowego wynosi

$$\text{var}(\varepsilon_i|X) = \text{var}(y_i - X_i\beta|X) = \text{var}(y_i) = X_i\beta(1 - X_i\beta)$$

```
      Source |       SS       df       MS              Number of obs =    6512
-------------+------------------------------           F(  6,  6505) =   35.09
       Model |  29.9105924     6  4.98509874            Prob > F      =  0.0000
    Residual |  924.099082  6505  .142059813            R-squared     =  0.0314
-------------+------------------------------           Adj R-squared =  0.0305
       Total |  954.009674  6511  .146522758            Root MSE      =  .37691
------------------------------------------------------------------------------
       praca |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |  -.0055054   .0007569    -7.27   0.000    -.0069891   -.0040217
   _Iwyksz_2 |   .0171293   .0153486     1.12   0.264    -.0129591    .0472177
   _Iwyksz_3 |  -.0939118   .0253777    -3.70   0.000    -.1436604   -.0441632
   _Iwyksz_4 |  -.0272158   .0140726    -1.93   0.053    -.0548027    .0003712
   _Iwyksz_5 |  -.1823766   .0191823    -9.51   0.000    -.2199803   -.1447729
        wies |   .0493767   .0097772     5.05   0.000     .0302102    .0685433
       _cons |   1.052533   .0316036    33.30   0.000     .9905798    1.114487
------------------------------------------------------------------------------

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         fit |      7008    .8223359     .067765   .5948861    .953877
```

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu normalnego

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu normalnego
- Zatem $Pr(y_i = 1) = \Phi(X_i\beta)$

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu normalnego
- Zatem $Pr(y_i = 1) = \Phi(X_i\beta)$
- Więc logarytm funkcji wiarogodności wynosi

$$\ell(\beta) = \sum_{i=1}^{N}(1 - y_i)\ln(1 - \Phi(X_i\beta)) + y_i \ln(\Phi(X\beta))$$

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu normalnego
- Zatem $Pr(y_i = 1) = \Phi(X_i\beta)$
- Więc logarytm funkcji wiarogodności wynosi

$$\ell(\beta) = \sum_{i=1}^{N}(1 - y_i)\ln(1 - \Phi(X_i\beta)) + y_i \ln(\Phi(X\beta))$$

- W celu identyfikacji parametrów modelu przyjmuje się, że $\sigma = 1$

```
Probit regression                                 Number of obs   =       6512
                                                  LR chi2(6)      =     192.41
                                                  Prob > chi2     =     0.0000
Log likelihood = -2956.5224                       Pseudo R2       =     0.0315
------------------------------------------------------------------------------
       praca |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |  -.0216518   .0030078    -7.20   0.000    -.0275469   -.0157567
   _Iwyksz_2 |   .0784266   .0634402     1.24   0.216    -.0459139    .2027671
   _Iwyksz_3 |  -.3488403   .0954592    -3.65   0.000    -.5359369   -.1617437
   _Iwyksz_4 |  -.1069245   .0567935    -1.88   0.060    -.2182378    .0043888
   _Iwyksz_5 |  -.6200677   .0717353    -8.64   0.000    -.7606664   -.4794691
        wies |   .1902053   .0387518     4.91   0.000     .1142532    .2661575
       _cons |   1.845056   .1278443    14.43   0.000     1.594486    2.095627
------------------------------------------------------------------------------
```

```
Marginal effects after probit
      y  = Pr(praca) (predict)
         =  .82867672
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.      z    P>|z|  [    95% C.I.   ]       X
---------+--------------------------------------------------------------------
    wiek |  -.0055063      .00076   -7.23   0.000     -.007 -.004013   40.3954
_Iwyks~2*|   .0195432      .01548    1.26   0.207  -.010797  .049884   .229883
_Iwyks~3*|  -.1017058      .03116   -3.26   0.001  -.162787 -.040625   .042998
_Iwyks~4*|  -.0272908      .01454   -1.88   0.061  -.055796  .001214   .463759
_Iwyks~5*|  -.1916381      .02543   -7.53   0.000   -.24149 -.141787   .103501
    wies*|   .0480969      .00973    4.94   0.000   .029031  .067162   .469287
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu logistycznego

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu logistycznego
- Zatem $P(X_i) = F(X_i\beta) = \frac{e^{-X_i\beta}}{1+e^{-X_i\beta}} = \Lambda(X_i\beta)$

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu logistycznego
- Zatem $P(X_i) = F(X_i\beta) = \frac{e^{-X_i\beta}}{1+e^{-X_i\beta}} = \Lambda(X_i\beta)$
- Więc logarytm funkcji wiarogodności wynosi

$$\ell(\beta) = \sum_{i=1}^{N}(1 - y_i)\ln(1 - \Lambda(X_i\beta)) + y_i \ln(\Lambda(X\beta))$$

- Zakładamy, że $F(\cdot)$ jest dystrybuantą rozkładu logistycznego
- Zatem $P(X_i) = F(X_i\beta) = \frac{e^{-X_i\beta}}{1+e^{-X_i\beta}} = \Lambda(X_i\beta)$
- Więc logarytm funkcji wiarogodności wynosi

$$\ell(\beta) = \sum_{i=1}^{N}(1 - y_i)\ln(1 - \Lambda(X_i\beta)) + y_i \ln(\Lambda(X\beta))$$

- Efekty cząstkowe są równe

$$\frac{\partial E(y = 1|X)}{\partial x_k} = \frac{\partial \Lambda(X_i\beta)}{\partial x_k} = \frac{e^{-X_i\beta}}{(1 + e^{-X_i\beta})^2} = \Lambda(X_i\beta)[1-\Lambda(X_i\beta)]\beta_k$$

- Inną zaleta modelu jest istnienie miary wpływu pojedynczej zmiennej na prawdopodobieństwo sukcesu, która nie zależy od wartości pozostałych zmiennych

- Inną zaleta modelu jest istnienie miary wpływu pojedynczej zmiennej na prawdopodobieństwo sukcesu, która nie zależy od wartości pozostałych zmiennych
- Iloraz szans (ang. *Odds-ratio*) definiujemy jako stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki

$$\mathrm{Odds}(x_i) = \frac{Pr(y = 1|x_i)}{Pr(y = 0|x_i)} = e^{x_i \beta}$$

- Inną zaleta modelu jest istnienie miary wpływu pojedynczej zmiennej na prawdopodobieństwo sukcesu, która nie zależy od wartości pozostałych zmiennych
- Iloraz szans (ang. *Odds-ratio*) definiujemy jako stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki

$$\mathrm{Odds}(x_i) = \frac{Pr(y = 1|x_i)}{Pr(y = 0|x_i)} = e^{x_i \beta}$$

- Prawdopodobieństwo wyrzucenia 6 oczek na sześciennej symetrycznej kostce wynosi $\frac{1}{6}$

- Inną zaleta modelu jest istnienie miary wpływu pojedynczej zmiennej na prawdopodobieństwo sukcesu, która nie zależy od wartości pozostałych zmiennych
- Iloraz szans (ang. *Odds-ratio*) definiujemy jako stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki

$$\mathrm{Odds}(x_i) = \frac{Pr(y = 1|x_i)}{Pr(y = 0|x_i)} = e^{x_i \beta}$$

- Prawdopodobieństwo wyrzucenia 6 oczek na sześciennej symetrycznej kostce wynosi $\frac{1}{6}$
- Szansa sukcesu wynosi 1 do 5

```
Logistic regression                               Number of obs   =       6512
                                                  LR chi2(6)      =     194.20
                                                  Prob > chi2     =     0.0000
Log likelihood = -2955.6251                       Pseudo R2       =     0.0318
------------------------------------------------------------------------------
       praca |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |  -.0391685   .0054266    -7.22   0.000    -.0498045   -.0285325
   _Iwyksz_2 |   .1433285   .1169148     1.23   0.220    -.0858204    .3724773
   _Iwyksz_3 |  -.6228615   .1659804    -3.75   0.000    -.9481772   -.2975458
   _Iwyksz_4 |  -.1997211   .1032286    -1.93   0.053    -.4020454    .0026032
   _Iwyksz_5 |  -1.080874   .1250762    -8.64   0.000    -1.326019   -.8357295
        wies |   .3492276   .0695027     5.02   0.000     .2130047    .4854504
       _cons |   3.205443   .2331851    13.75   0.000     2.748409    3.662477
------------------------------------------------------------------------------
```

```
Logistic regression                                     Number of obs   =       6512
                                                        LR chi2(6)      =     194.20
                                                        Prob > chi2     =     0.0000
Log likelihood = -2955.6251                             Pseudo R2       =     0.0318
------------------------------------------------------------------------------
       praca | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |   .9615887   .0052182    -7.22   0.000     .9514154    .9718707
   _Iwyksz_2 |   1.154109   .1349324     1.23   0.220     .9177591    1.451326
   _Iwyksz_3 |   .5364073   .0890331    -3.75   0.000     .3874466    .7426386
   _Iwyksz_4 |   .8189591     .08454    -1.93   0.053     .6689504    1.002607
   _Iwyksz_5 |   .3392987   .0424382    -8.64   0.000     .2655321    .4335581
        wies |   1.417972   .0985529     5.02   0.000     1.237391    1.624907
------------------------------------------------------------------------------
```

Porównanie modeli

- Wszystkie trzy modele dla zmiennej binarnej dają takie same oszacowania wielkości efektów cząstkowych dla wartości średnich zmiennych niezależnych

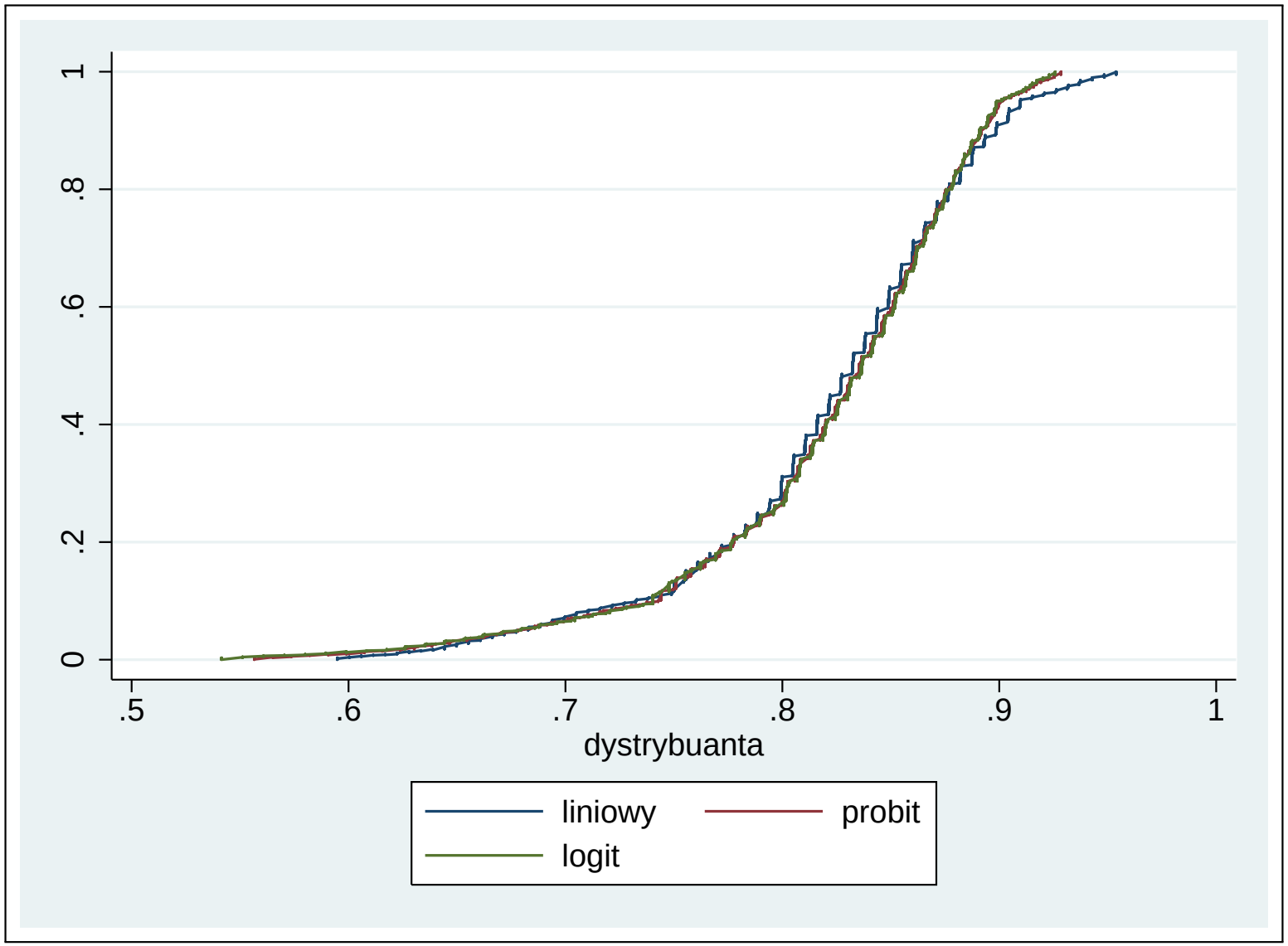
Porównanie modeli

- Wszystkie trzy modele dla zmiennej binarnej dają takie same oszacowania wielkości efektów cząstkowych dla wartości średnich zmiennych niezależnych
- Modele różnią się wielkościami oszacowań parametrów

Porównanie modeli

- Wszystkie trzy modele dla zmiennej binarnej dają takie same oszacowania wielkości efektów cząstkowych dla wartości średnich zmiennych niezależnych
- Modele różnią się wielkościami oszacowań parametrów
- Przybliżone relacje między parametrami uzyskujemy korzystając z symetryczności rozkładów, więc $\bar{X}_i\beta = 0$, zatem
  - $f_{LMP}(0) = 1$
  - $f_{PR}(0) = \frac{1}{\sqrt{2\Pi}} \approx 0{,}4$
  - $f_L(0) = 0{,}25$

Variable	reg	prob	log
wiek	-.00550541***	-.02165179***	-.03916851***
_Iwyksz_2	.01712929	.07842655	.14332849
_Iwyksz_3	-.09391179***	-.34884031***	-.6228615***
_Iwyksz_4	-.02721575	-.1069245	-.19972112
_Iwyksz_5	-.18237663***	-.62006774***	-1.0808745***
wies	.04937674***	.19020531***	.34922757***
_cons	1.0525333***	1.8450564***	3.2054431***


0
.2
.4
.6
.8
1
.5
.6
.7
.8
.9
1
dystrybuanta
liniowy
probit
logit

- Dla modeli z binarną zmienną zależną można jako miary dopasowania użyć pseudo-R^2 Mc-Faddena

- Dla modeli z binarną zmienną zależną można jako miary dopasowania użyć pseudo-R^2 Mc-Faddena
- Można jednak pseudo-R^2 zdefiniować inaczej

- Dla modeli z binarną zmienną zależną można jako miary dopasowania użyć pseudo-R^2 Mc-Faddena
- Można jednak pseudo-R^2 zdefiniować inaczej
- W KMRL R^2 można przedstawić jako

$$R^2 = \frac{ESS}{ESS + RSS} = \frac{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2}{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2 + \tilde{\sigma}^2}$$

gdzie $\tilde{\sigma}^2$ to estymator MNW wariancji

- Dla modeli z binarną zmienną zależną można jako miary dopasowania użyć pseudo-R^2 Mc-Faddena
- Można jednak pseudo-R^2 zdefiniować inaczej
- W KMRL R^2 można przedstawić jako

$$R^2 = \frac{ESS}{ESS + RSS} = \frac{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2}{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2 + \tilde{\sigma}^2}$$

  gdzie $\tilde{\sigma}^2$ to estymator MNW wariancji
- Tę zależność wykorzystali Mc-Kelvey i Zavoina definiując pseudo-R^2 dla zmiennej ukrytej

$$R^2_{MZ} = \frac{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2}{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2 + \tilde{\sigma}^2}$$

- Dla modeli z binarną zmienną zależną można jako miary dopasowania użyć pseudo-R^2 Mc-Faddena
- Można jednak pseudo-R^2 zdefiniować inaczej
- W KMRL R^2 można przedstawić jako

$$R^2 = \frac{ESS}{ESS + RSS} = \frac{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2}{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2 + \tilde{\sigma}^2}$$

  gdzie $\tilde{\sigma}^2$ to estymator MNW wariancji
- Tę zależność wykorzystali Mc-Kelvey i Zavoina definiując pseudo-R^2 dla zmiennej ukrytej

$$R^2_{MZ} = \frac{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2}{\frac{1}{N}\sum(\hat{y}_i - \bar{y})^2 + \tilde{\sigma}^2}$$

- Opisuje ono procent wyjaśnienia wariancji zmiennej ukrytej, gdyby była ona bezpośrednio obserwowana

- Alternatywą jest badanie właściwości predykcyjnych modelu

Prognozowane	0	1	
	$p < \bar{p}$	$p < \bar{p}$	
0	n_{00}	n_{01}	$n_{00} + n_{01}$
1	n_{10}	n_{11}	$n_{10} + n_{11}$
Suma	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n

- Alternatywą jest badanie właściwości predykcyjnych modelu
- Zakładamy, że model przewiduje sukces, gdy $\hat{p} > \bar{p}$; $\bar{p}$ nazywamy wartością progową.

Prognozowane	0	1	
	$p < \bar{p}$	$p < \bar{p}$	
0	n_{00}	n_{01}	$n_{00} + n_{01}$
1	n_{10}	n_{11}	$n_{10} + n_{11}$
Suma	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n

- Na podstawie tablicy klasyfikacyjnej można obliczyć R^2 liczebnościowe

$$R^2_{licz} = \frac{n_{00} + n_{11}}{n}$$

- Na podstawie tablicy klasyfikacyjnej można obliczyć R^2 liczebnościowe

$$R^2_{licz} = \frac{n_{00} + n_{11}}{n}$$

- Losowo przydzielając jako wartości prognozowane 0 i 1 uzyskamy $R^2_{licz} = \frac{1}{2}$

- Na podstawie tablicy klasyfikacyjnej można obliczyć R^2 liczebnościowe

$$R^2_{licz} = \frac{n_{00} + n_{11}}{n}$$

- Losowo przydzielając jako wartości prognozowane 0 i 1 uzyskamy $R^2_{licz} = \frac{1}{2}$
- Skorygowane R^2 liczebnościowe

$$\bar{R}^2_{licz} = \frac{\sum_j n_{jj} - n_{max}}{n - n_{max}}$$

gdzie $n_{max} = \max(n_{00}, n_{01}, n_{10}, n_{11})$

- Na podstawie tablicy klasyfikacyjnej można obliczyć R^2 liczebnościowe

$$R^2_{licz} = \frac{n_{00} + n_{11}}{n}$$

- Losowo przydzielając jako wartości prognozowane 0 i 1 uzyskamy $R^2_{licz} = \frac{1}{2}$
- Skorygowane R^2 liczebnościowe

$$\bar{R}^2_{licz} = \frac{\sum_j n_{jj} - n_{max}}{n - n_{max}}$$

gdzie $n_{max} = \max(n_{00}, n_{01}, n_{10}, n_{11})$

- $\bar{R}^2_{licz} \leq 1$, ale $\bar{R}^2_{licz} > 0$, gdy udział prawidłowo przewidzianych zdarzeń przekracza udział najczęstszego zdarzenia w próbie

Wrażliwość i specyficzność

- Na podstawie wartości z tablicy klasyfikacyjnej można zdefiniować dwie miary

Wrażliwość i specyficzność

- Na podstawie wartości z tablicy klasyfikacyjnej można zdefiniować dwie miary
- Wrażliwość

$$Pr(\hat{p}_i > \bar{p} | y_i = 1) \approx \frac{n_{11}}{n_{01} + n_{11}}$$

Wrażliwość i specyficzność

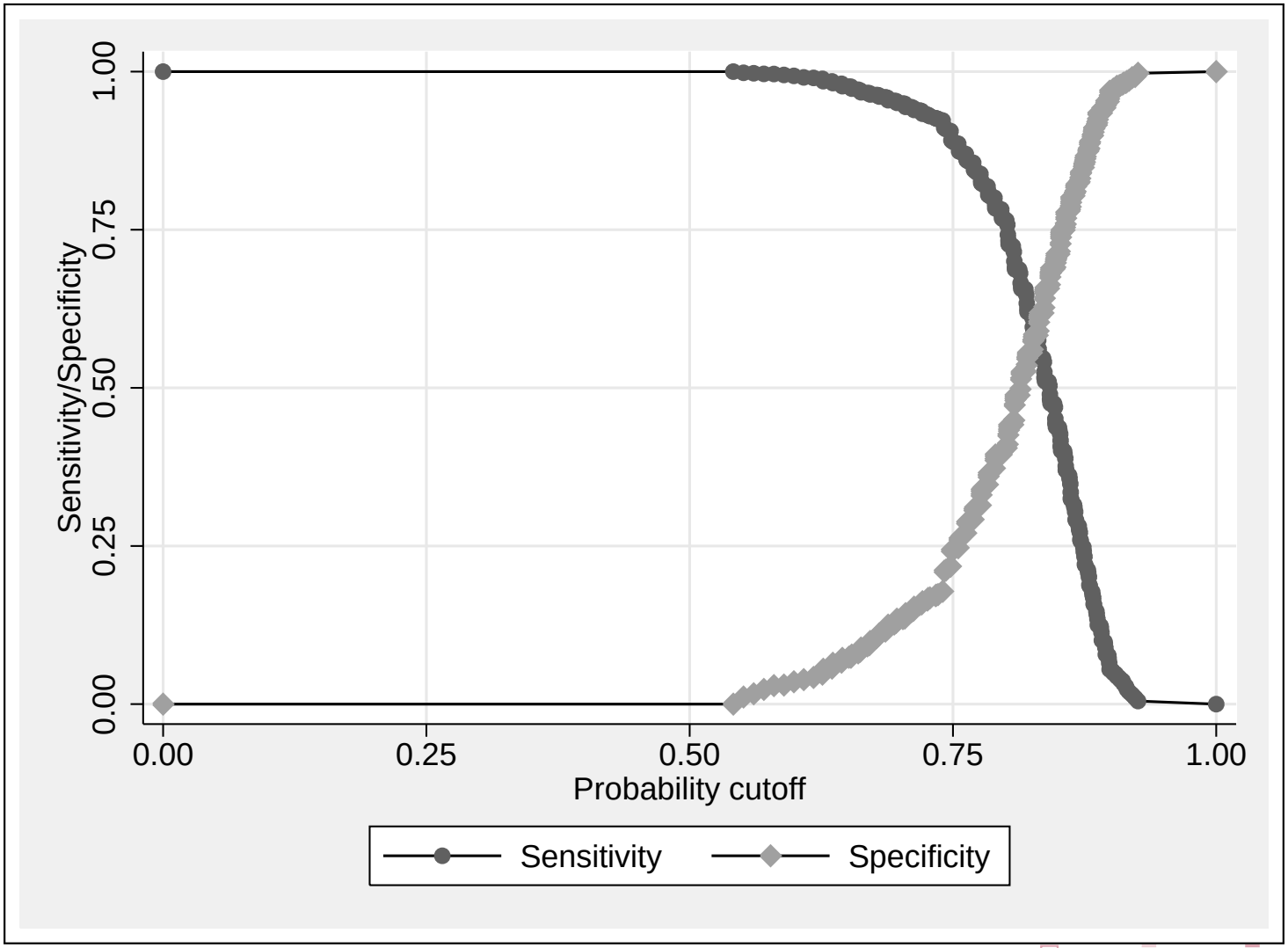
- Na podstawie wartości z tablicy klasyfikacyjnej można zdefiniować dwie miary
- Wrażliwość

$$Pr(\hat{p}_i > \bar{p} | y_i = 1) \approx \frac{n_{11}}{n_{01} + n_{11}}$$

- Specyficzność

$$Pr(\hat{p}_i < \bar{p} | y_i = 0) \approx \frac{n_{00}}{n_{00} + n_{10}}$$

Wrażliwość i specyficzność



Wrażliwość i specyficzność

```
Logistic model for praca
              -------- True --------
Classified |         D             ~D  |      Total
-----------+--------------------------+-----------
     +     |      4093            691  |       4784
     -     |      1258            470  |       1728
-----------+--------------------------+-----------
   Total   |      5351           1161  |       6512

Classified + if predicted Pr(D) >= .8
True D defined as praca != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   76.49%
Specificity                     Pr( -|~D)   40.48%
Positive predictive value       Pr( D| +)   85.56%
Negative predictive value       Pr(~D| -)   27.20%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   59.52%
False - rate for true D         Pr( -| D)   23.51%
False + rate for classified +   Pr(~D| +)   14.44%
False - rate for classified -   Pr( D| -)   72.80%
```

Krzywa ROC

- Fałszywe przewidywania

$$FP = Pr(\hat{p}_i > \bar{p} | y_i = 0)$$

Krzywa ROC

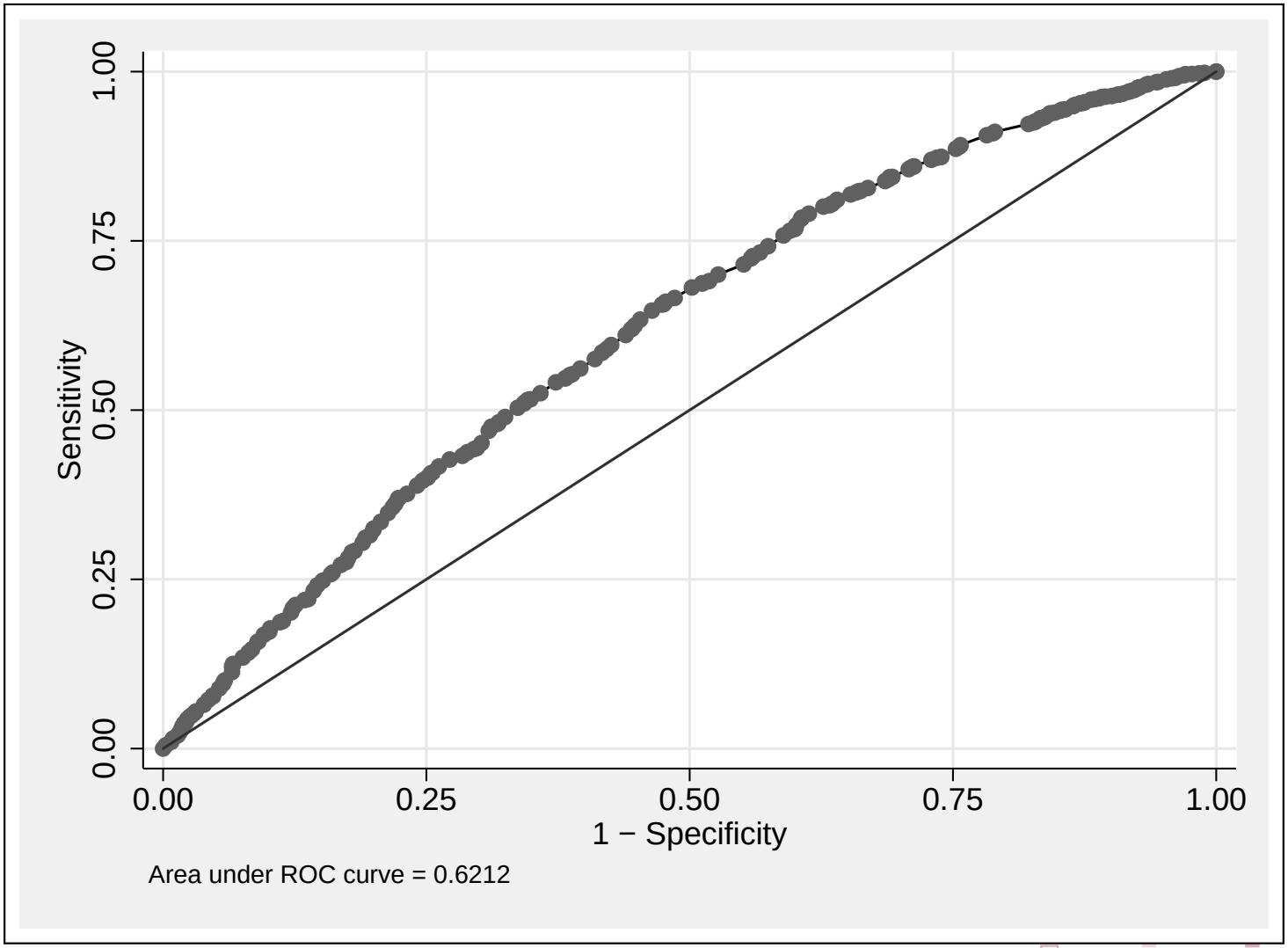
- Fałszywe przewidywania

$$FP = Pr(\hat{p}_i > \bar{p} | y_i = 0)$$

- Fałszywe przewidywania i specyficzność łączy zależność

$$FP = Pr(\hat{p}_i > \bar{p} | y_i = 0) = 1 - Pr(\hat{p}_i < \bar{p} | y_i = 0) = 1 - \text{specyficzność}$$

Krzywa ROC


Sensitivity
0.00
0.25
0.50
0.75
1.00
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Area under ROC curve = 0.6212

- Odpowiednikiem testu RESET jest test typu związku (linktest)

```
Logistic regression                               Number of obs   =       6512
                                                  LR chi2(2)      =     197.12
                                                  Prob > chi2     =     0.0000
Log likelihood = -2954.1693                       Pseudo R2       =     0.0323
------------------------------------------------------------------------------
       praca |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   1.537635   .3218281     4.78   0.000     .9068632    2.168406
      _hatsq |   -.197935   .1152741    -1.72   0.086     -.423868    .0279981
       _cons |   -.320369   .2171085    -1.48   0.140    -.7458938    .1051558
------------------------------------------------------------------------------
```

- Test jakości dopasowania (gof) polega na podziale próby na podpróbki i badaniu oszacowania stałej

- Test jakości dopasowania (gof) polega na podziale próby na podpróbki i badaniu oszacowania stałej
- Statystyka Hosmera i Lemeshowa

$$LM = \sum_{r=1}^{R-1} \frac{n_r(p - \hat{p})^2}{(\hat{p}(1 - \hat{p}))} \sim \chi^2(R)$$

```
Logistic model for praca, goodness-of-fit test

      number of observations =      6512
 number of covariate patterns =       209
           Pearson chi2(202) =    201.74
                 Prob > chi2 =    0.4918
```