

Binarne zmienne zależne

cz. II

Stanisław Cichocki

Natalia Nehrebecka

Plan zajęć

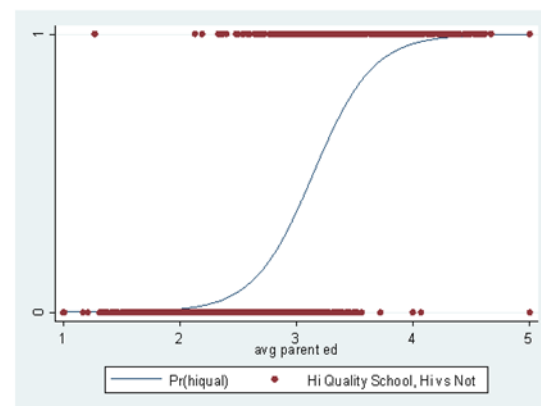
1. Wstęp
 - a) Binarne zmienne zależne
 - b) Interpretacja ekonomiczna
 - c) Interpretacja współczynników
2. Liniowy model prawdopodobieństwa
 - a) Interpretacja współczynników
3. **Probit**
 - a) Interpretacja współczynników
 - b) Miary dopasowania
 - c) Diagnostyka
4. Logit
 - a) Interpretacja współczynników
 - b) Miary dopasowania
 - c) Diagnostyka

Probit

- ▶ W modelu **probitowym** zakładamy, że $F()$ jest dystrybuantą rozkładu normalnego
- ▶ Założenia modelu probitowego:
 - Obserwacje są niezależne
 - Rozkład warunkowy:

$$\begin{aligned} Pr(y_i|x_i) &= \begin{cases} 1 - \Phi(x_i\beta) & \text{dla } y_i = 0 \\ \Phi(x_i\beta) & \text{dla } y_i = 1 \end{cases} \\ &= [1 - \Phi(x_i\beta)]^{1-y_i} \Phi(x_i\beta)^{y_i} \end{aligned}$$

- Gdzie Φ oznacza dystrybuantę rozkładu normalnego $N(0,1)$



Probit

- ▶ Funkcja wiarygodności ma następującą postać:

$$L(\beta) = \prod_{i=1}^n [1 - \Phi(x_i\beta)]^{1-y_i} \Phi(x_i\beta)^{y_i}$$

- ▶ Logarytm funkcji wiarygodności będzie miał postać:

$$l(\beta) = \sum_{i=1}^n [(1 - y_i) \ln(1 - \Phi(x_i\beta)) + y_i \ln \Phi(x_i\beta)]$$

Probit

- ▶ Wartość oczekiwana:

$$E(y_i|x_i) = 1 \cdot \Phi(x_i\beta) + 0 \cdot (1 - \Phi(x_i\beta)) = \Phi(x_i\beta)$$

- ▶ Efekt cząstkowy dla zmiennej x_k :

$$\frac{\partial E(y|x)}{\partial x_k} = \frac{\partial \Phi(x_i\beta)}{\partial x_k} = \phi(\mathbf{x}_i\beta)\beta_k$$

Probit - Interpretacja współczynników

- ▶ nie interpretuje się współczynników w modelu probitowym
- ▶ interpretuje się efekty cząstkowe (*krańcowe*):
 - a) dla zmiennych objaśniających ciągłych:
 - wpływ jednostkowej zmiany zmiennej niezależnej na wielkość prawdopodobieństwa sukcesu;
 - efekty cząstkowe dla zmiennych objaśniających ciągłych liczymy zwykle dla średnich wartości tych zmiennych (efekty cząstkowe zależą od wielkości zmiennych objaśniających)
 - b) dla zmiennych objaśniających zero-jedynkowych:
 - różnica między prawdopodobieństwem sukcesu dla zmiennej zero-jedynkowej równej 0 i równej 1, przy pozostałych zmiennych ustalonych na poziomie średnich

Probit - Interpretacja współczynników

- ▶ znak efektu cząstkowego dla danej zmiennej jest taki sam jak znak współczynnika przy tej zmiennej
- ▶ możemy zatem interpretować znaki przy współczynnikach:
 - dodatni znak \Rightarrow zmienna wpływa dodatnio na prawdopodobieństwo sukcesu
 - ujemny znak \Rightarrow zmienna wpływa ujemnie na prawdopodobieństwo sukcesu

Probit

Probit regression

Log likelihood = -2881.7371

Number of obs = 4,877
LR chi2(11) = 322.58
Prob > chi2 = 0.0000
Pseudo R2 = 0.0530

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
stateur	.0566916	.0094341	6.01	0.000	.0382012	.075182
statemb	.003684	.0006012	6.13	0.000	.0025057	.0048624
age	.0117275	.0021911	5.35	0.000	.0074331	.0160219
tenure 	.0173067	.0037642	4.60	0.000	.009929	.0246844
slack	.3661648	.0395464	9.26	0.000	.2886553	.4436743
male	-.1139709	.0519568	-2.19	0.028	-.2158042	-.0121375
smsa	-.0974368	.041559	-2.34	0.019	-.178891	-.0159826
married	.1441728	.0415553	3.47	0.001	.0627259	.2256198
yrdispl	-.0371914	.0090398	-4.11	0.000	-.0549092	-.0194736
rr2	-.7020753	.2523342	-2.78	0.005	-1.196641	-.2075092
head 	-.1101841	.0473191	-2.33	0.020	-.2029279	-.0174403
_cons	-.8093586	.1530845	-5.29	0.000	-1.109399	-.5093185

Probit – Interpretacja efektów cząstkowych (krańcowych)

Marginal effects after probit

y = Pr(y) (predict)

= .69402302

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
stateur	.019886	.0033	6.02	0.000	.013409	.026363		7.51103
statemb	.0012923	.00021	6.13	0.000	.000879	.001705		180.66
age	.0041137	.00077	5.36	0.000	.002608	.005619		36.13
tenure	.0060708	.00132	4.60	0.000	.003485	.008657		5.66414
slack*	.1273458	.01356	9.39	0.000	.100772	.15392		.476112
male*	-.0393392	.01763	-2.23	0.026	-.073892	-.004786		.764199
smsa*	-.033908	.01434	-2.36	0.018	-.062018	-.005798		.652655
married*	.0510228	.01482	3.44	0.001	.021973	.080073		.632766
yrdispl	-.0130458	.00317	-4.12	0.000	-.019259	-.006832		5.20361
rr2	-.246271	.08851	-2.78	0.005	-.419752	-.07279		.20344
head*	-.0382403	.01624	-2.36	0.019	-.070064	-.006417		.680541

Probit – Interpretacja efektów cząstkowych (krańcowych)

- ▶ Efekty cząstkowe dla mężczyzny, w wieku 35 lat, który jest żonaty i jest głową rodziny;
- ▶ reszta zmiennych na poziomie średniej z próby

```
mfx, at(male = 1 age = 35 married = 1 head = 1)
```

```
Marginal effects after probit
```

```
    y  = Pr(y) (predict)  
      = .68612838
```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
stateur	.0201081	.00334	6.02	0.000	.013557 .02666	7.51103
statemb	.0013067	.00021	6.09	0.000	.000886 .001727	180.66
age	.0041597	.00079	5.28	0.000	.002614 .005705	35
tenure	.0061386	.00133	4.60	0.000	.003525 .008752	5.66414
slack*	.1287766	.01379	9.34	0.000	.101746 .155807	.476112
male*	-.039244	.01744	-2.25	0.024	-.073428 -.00506	1
smsa*	-.0342977	.01454	-2.36	0.018	-.062802 -.005794	.652655
married*	.0527807	.01536	3.44	0.001	.022675 .082886	1
yrdispl	-.0131915	.00322	-4.10	0.000	-.019499 -.006884	5.20361
rr2	-.2490205	.09009	-2.76	0.006	-.425593 -.072448	.20344
head*	-.0379799	.01599	-2.38	0.018	-.069321 -.006639	1

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

Probit – Interpretacja efektów cząstkowych (krańcowych)

- ▶ Efekty cząstkowe dla kobiety w wieku 35 lat, która nie jest zamężna i nie jest głową rodziny;
- ▶ reszta zmiennych na poziomie średniej z próby

```
mfx, at(male = 0 age = 35 married = 0 head = 0)
```

```
Marginal effects after probit
```

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

```
      = .71392498
```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
stateur	.0192814	.00325	5.93	0.000	.01291 .025653	7.51103
statemb	.001253	.0002	6.16	0.000	.000854 .001652	180.66
age	.0039886	.00075	5.32	0.000	.002519 .005458	35
tenure	.0058862	.00128	4.59	0.000	.003375 .008398	5.66414
slack*	.1234544	.01338	9.23	0.000	.097227 .149681	.476112
male*	-.0399497	.0182	-2.19	0.028	-.075629 -.004271	0
smsa*	-.0328496	.01385	-2.37	0.018	-.059993 -.005706	.652655
married*	.0469316	.01354	3.47	0.001	.020403 .07346	0
yrdispl	-.0126492	.00307	-4.12	0.000	-.018664 -.006634	5.20361
rr2	-.2387825	.08433	-2.83	0.005	-.404059 -.073506	.20344
head*	-.0385862	.0168	-2.30	0.022	-.071517 -.005656	0

Plan zajęć

1. Wstęp
 - a) Binarne zmienne zależne
 - b) Interpretacja ekonomiczna
 - c) Interpretacja współczynników
2. Liniowy model prawdopodobieństwa
 - a) Interpretacja współczynników
3. Probit
 - a) Interpretacja współczynników
 - b) Miary dopasowania
 - c) Diagnostyka
4. Logit
 - a) Interpretacja współczynników
 - b) Miary dopasowania
 - c) Diagnostyka

Miary dopasowania

- ▶ Często stosowaną miarą dopasowania jest:

$$R_{McFadden}^2 = 1 - \frac{l(\beta)}{l(\beta_R)}$$

- gdzie β_R zawiera jedynie stałą
- ▶ *pseudo* – R^2 spełnia

$$0 \leq R_{McFadden}^2 \leq 1$$

- ▶ *pseudo* – R^2 nie można jednak w pełni ściśle interpretować jako procent zmienności wyjaśnionej przez model.

Probit

```
probit y stateur statemb age tenure slack male smsa married yrdispl rr2 head
```

```
Iteration 0: log likelihood = -3043.028
Iteration 1: log likelihood = -2882.6368
Iteration 2: log likelihood = -2881.7372
Iteration 3: log likelihood = -2881.7371
```

Probit regression

```
Number of obs   =      4877
LR chi2(11)     =      322.58
Prob > chi2     =      0.0000
Pseudo R2      =      0.0530
```

Log likelihood = -2881.7371

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
stateur	.0566916	.0094341	6.01	0.000	.0382012	.075182
statemb	.003684	.0006012	6.13	0.000	.0025057	.0048624
age	.0117275	.0021911	5.35	0.000	.0074331	.0160219
tenure	.0173067	.0037642	4.60	0.000	.009929	.0246844
slack	.3661648	.0395464	9.26	0.000	.2886553	.4436743
male	-.1139709	.0519568	-2.19	0.028	-.2158042	-.0121375
smsa	-.0974368	.041559	-2.34	0.019	-.178891	-.0159826
married	.1441728	.0415553	3.47	0.001	.0627259	.2256198
yrdispl	-.0371914	.0090398	-4.11	0.000	-.0549092	-.0194736
rr2	-.7020753	.2523342	-2.78	0.005	-1.196641	-.2075092
head	-.1101841	.0473191	-2.33	0.020	-.2029279	-.0174403
_cons	-.8093586	.1530845	-5.29	0.000	-1.109399	-.5093185

Miary dopasowania

- ▶ W przypadku *pseudo* – R^2 istnieje problem związany z tym, iż miara ta zawsze rośnie wraz z dodawaniem zmiennych do modelu.

$$\bar{R}_{McFadden}^2 = 1 - \frac{l(\beta) - K}{l(\beta_R)}$$

- Gdzie
- β_R zawiera jedynie stałą;
- K jest liczbą parametrów w modelu.

Miary dopasowania

Measures of Fit for probit of y

Log-Lik Intercept Only:	-3043.028	Log-Lik Full Model:	-2881.737
D(4865):	5763.474	LR(11):	322.582
		Prob > LR:	0.000
McFadden's R2:	0.053	McFadden's Adj R2:	0.049
Maximum Likelihood R2:	0.064	Cragg & Uhler's R2:	0.090
McKelvey and Zavoina's R2:	0.110	Efron's R2:	0.067
Variance of y*:	1.124	Variance of error:	1.000
Count R2:	0.699	Adj Count R2:	0.048

Miary dopasowania

- ▶ Dla modeli dla zmiennych binarnych zależnych można zdefiniować *pseudo* – R^2 inne niż *pseudo* – $R_{McFadden}^2$, które są bliższe definicji R^2 w MNK.
- ▶ W KMRL R^2 można zapisać jako:

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

- gdzie $\tilde{\sigma}^2$ jest estymatorem MNW wariancji składnika losowego
- $\tilde{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N e_i^2$

Miary dopasowania

- ▶ Analogiczną miarę dopasowania można zdefiniować dla binarnej zmiennej zależnej.
- ▶ *pseudo* – R^2 będzie mierzyło udział wariancji wartości dopasowanych w całkowitej wariancji *zmiennej ukrytej*.

$$R_{McKelvey, Zavoina}^2 = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i^* - \bar{\hat{y}}^*)^2}{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i^* - \bar{\hat{y}}^*)^2 + \tilde{\sigma}^2}$$

- Gdzie:
- $\tilde{\sigma}^2$ jest estymatorem *MNW* wariancji składnika losowego;
- $\hat{y}_i^* = x\tilde{\beta}$ jest oszacowaniem z modelu dla wielkości zmiennej ukrytej.
- ▶ *pseudo* – R^2 McKelveya i Zavoiny **opisuje procent wyjaśnienia, jaki uzyskaliśmy w modelu dla zmiennej ukrytej y_i^* , gdyby była ona bezpośrednio obserwowalna.**

Miary dopasowania

- ▶ W modelu probitowym

$$R_{McKelvey, Zavoina}^2 = \frac{ESS}{ESS+TSS} = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i^* - \bar{\hat{y}}^*)^2}{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i^* - \bar{\hat{y}}^*)^2 + \mathbf{1}}$$

- ▶ gdzie:
 - N - liczba obserwacji w próbie,
 - $\bar{\hat{y}}^*$ - średnia wartość dopasowana zmiennej ukrytej y_i^* .

Miary dopasowania

Measures of Fit for probit of y

Log-Lik Intercept Only:	-3043.028	Log-Lik Full Model:	-2881.737
D(4865):	5763.474	LR(11):	322.582
		Prob > LR:	0.000
McFadden's R2:	0.053	McFadden's Adj R2:	0.049
Maximum Likelihood R2:	0.064	Cragg & Uhler's R2:	0.090
McKelvey and Zavoina's R2:	0.110	Efron's R2:	0.067
Variance of y*:	1.124	Variance of error:	1.000
Count R2:	0.699	Adj Count R2:	0.048

Miary dopasowania

- ▶ tablica klasyfikacyjna:

		<i>Zaobserwowane</i>		
<i>Prognozowane</i>		0	1	Razem
	0	n_{00}	n_{01}	$n_{00} + n_{01}$
	1	n_{10}	n_{11}	$n_{10} + n_{11}$
	Razem	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n

Miary dopasowania

- ▶ standardowo punktem granicznym jest $p^* = 0.5$ dla poprawnie zakwalifikowanego wyniku

Probit model for y						
----- True -----						
Classified		D ~D		Total		
-----+-----+-----						
+		3197		1330		4527
-		138		212		350
-----+-----+-----						
Total		3335		1542		4877

Classified + if predicted $\Pr(D) \geq .5$

Miary dopasowania

- ▶ na podstawie tablicy klasyfikacyjnej można ustalić na ile model trafnie przewiduje sukcesy i porażki
- ▶ *liczebnościowe* R^2 :

$$R^2_{liczebnościowe} = \frac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{10} + n_{01}}$$

Miary dopasowania

- ▶ wielkość *liczebnościowego* R^2 może być myląca ➡
w przypadku zmiennej binarnej można zawsze uzyskać **co najmniej 50%** trafności przewidywań prognozując dla wszystkich obserwacji wartość y_i , która jest najczęściej obserwowana w próbie
- ▶ *skorygowane liczebnościowe* R^2 :

$$\bar{R}_{\text{liczebnościowe}}^2 = \frac{n_{00} + n_{11} - n_{\max}}{n_{00} + n_{11} + n_{10} + n_{01} - n_{\max}}$$

- ▶ Gdzie:
 - n_{\max} - liczba obserwacji dla zdarzenia (*sukces/porażka*), którą obserwowano częściej

Miary dopasowania

- ▶ *skorygowane liczebnościowe R^2*
 - **zawsze ≤ 1** , ale
 - **> 0** tylko wtedy, gdy udział prawidłowo przewidzianych zdarzeń przekracza udział zdarzenia, które było najczęściej obserwowane w próbie
- ▶ *skorygowane liczebnościowe R^2* : jest to „procent” prawidłowych przewidywań uzyskany poprzez zastosowanie w modelu zmiennych objaśniających

Miary dopasowania

Measures of Fit for probit of y

Log-Lik Intercept Only:	-3043.028	Log-Lik Full Model:	-2881.737
D(4865):	5763.474	LR(11):	322.582
		Prob > LR:	0.000
McFadden's R2:	0.053	McFadden's Adj R2:	0.049
Maximum Likelihood R2:	0.064	Cragg & Uhler's R2:	0.090
McKelvey and Zavoina's R2:	0.110	Efron's R2:	0.067
Variance of y*:	1.124	Variance of error:	1.000
Count R2:	0.699	Adj Count R2:	0.048

Miary dopasowania

- ▶ tablica klasyfikacyjna może być użyta do zdefiniowania jeszcze 2 miar:
 - a) **wrażliwość** – prawdopodobieństwo przewidzenia sukcesu dla obserwacji, dla której zaobserwowano sukces

$$\text{wrażliwość} = \Pr(\hat{p}_i \geq p^* | y_i = 1) \approx \frac{n_{11}}{n_{01} + n_{11}}$$

Miary dopasowania

- b) **specyficzność** – prawdopodobieństwo przewidzenia porażki dla obserwacji, dla której zaobserwowano porażkę

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

Miary dopasowania

Probit model for y

		----- True -----		
Classified		D	~D	Total
-----+-----+-----+-----				
+		3197	1330	4527
-		138	212	350
-----+-----+-----+-----				
Total		3335	1542	4877

Classified + if predicted $\Pr(D) \geq .5$

True D defined as y != 0

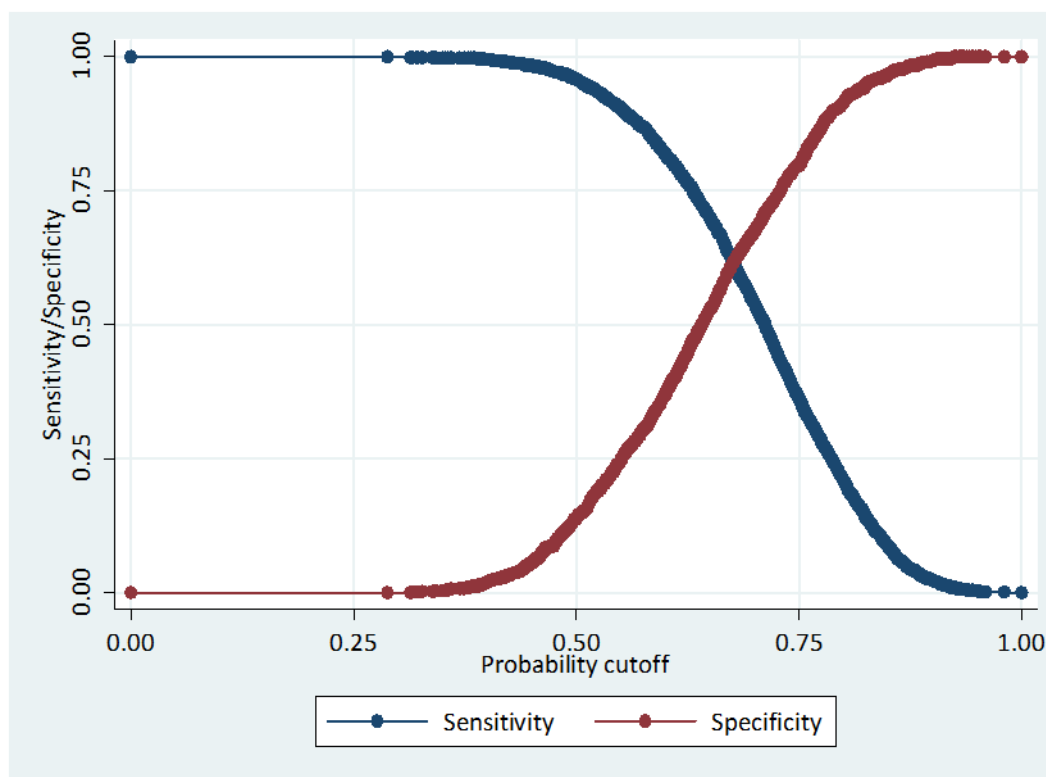
Sensitivity	Pr(+ D)	95.86%
Specificity	Pr(- ~D)	13.75%
Positive predictive value	Pr(D +)	70.62%
Negative predictive value	Pr(~D -)	60.57%

False + rate for true ~D	Pr(+ ~D)	86.25%
False - rate for true D	Pr(- D)	4.14%
False + rate for classified +	Pr(~D +)	29.38%
False - rate for classified -	Pr(D -)	39.43%

Correctly classified		69.90%

Miary dopasowania

- ▶ wrażliwość i specyficzność



Miary dopasowania

Probit model for y

Classified	----- True -----		Total
	D	~D	
+	2015	577	2592
-	1320	965	2285
Total	3335	1542	4877

Classified + if predicted $\Pr(D) \geq .68$

True D defined as $y \neq 0$

Sensitivity	Pr(+ D)	60.42%
Specificity	Pr(- ~D)	62.58%
Positive predictive value	Pr(D +)	77.74%
Negative predictive value	Pr(~D -)	42.23%
False + rate for true ~D	Pr(+ ~D)	37.42%
False - rate for true D	Pr(- D)	39.58%
False + rate for classified +	Pr(~D +)	22.26%
False - rate for classified -	Pr(D -)	57.77%
Correctly classified		61.10%

Miary dopasowania

▶ dodatkowo:

c) **1- wrażliwość** – prawdopodobieństwo przewidzenia porażki dla obserwacji, dla której **zaobserwowano sukces**

$$1 - \text{wrażliwość} = \frac{n_{01}}{n_{01} + n_{11}}$$

Miary dopasowania

- d) **1- specyficzność** – prawdopodobieństwo przewidzenia sukcesu dla obserwacji, dla której **zaobserwowano porażkę**

$$1 - \text{specyficzność} = \frac{n_{10}}{n_{00} + n_{10}}$$

ROC (Receiver operating characteristic)

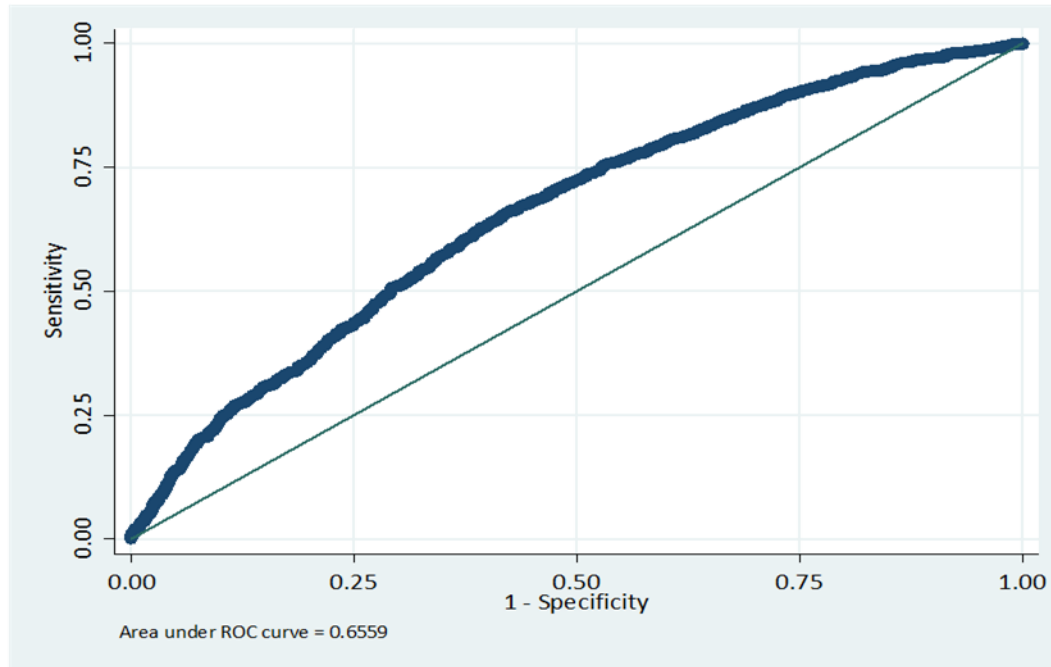
- ▶ Krzywa *ROC* obrazuje zależność pomiędzy wrażliwością (*TR* – *True Positives*)

$$TR = \text{wrażliwość} = \Pr(\hat{p}_i \geq p^* | y_i = 1)$$

- ▶ i prawdopodobieństwem uzyskania fałszywych przewidywanych sukcesów (*FP* – *False Positives*)

$$FP = 1 - \text{specyficzność} = \Pr(\hat{p}_i \geq p^* | y_i = 0)$$

ROC (*Receiver operating characteristic*)



- ▶ Im lepiej nasz model przewiduje, tym bardziej krzywa *ROC* odgięta jest w kierunku górnego rogu rysunku.
- ▶ Pole pod krzywą używane jest jako miara jakości dopasowania modelu.
 - $AUROC = 0.6559$ (max 1)

Plan zajęć

1. Wstęp
 - a) Binarne zmienne zależne
 - b) Interpretacja ekonomiczna
 - c) Interpretacja współczynników
2. Liniowy model prawdopodobieństwa
 - a) Interpretacja współczynników
3. Probit
 - a) Interpretacja współczynników
 - b) Miary dopasowania
 - c) Diagnostyka
4. Logit
 - a) Interpretacja współczynników
 - b) Miary dopasowania

Diagnostyka

▶ Testy na poprawność formy funkcyjnej

- W przypadku modeli z binarną zmienną objaśnianą zastosowanie znajduje *linktest*, który jest uogólnieniem i jednocześnie słabszą wersją testu *RESET*.
- Test ten polega na przeprowadzeniu modelu probitowego y_i na stałą, \hat{y}_i^* oraz $(\hat{y}_i^*)^2$.
- Istotny współczynnik przy $(\hat{y}_i^*)^2$ powoduje odrzucenie hipotezy o poprawności formy funkcyjnej modelu.

Diagnostyka

► Testy na poprawność formy funkcyjnej

```
Iteration 0:    log likelihood =  -3043.028
Iteration 1:    log likelihood = -2879.6536
Iteration 2:    log likelihood = -2879.0342
Iteration 3:    log likelihood = -2879.0339
Iteration 4:    log likelihood = -2879.0339
```

Probit regression

```
Number of obs   =      4877
LR chi2(2)      =      327.99
Prob > chi2     =      0.0000
Pseudo R2      =      0.0539
```

Log likelihood = -2879.0339

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
_hat	1.28044	.1331964	9.61	0.000	1.01938	1.5415
_hatsq 	-.2822587	.1200489	-2.35	0.019	-.5175501	-.0469672
_cons	-.0374954	.0365824	-1.02	0.305	-.1091955	.0342047
-----+-----						

Pytania teoretyczne

1. Wyjaśnić, czym jest wrażliwość i specyficzność modelu dla zmiennej binarnej i jak one zależą od progowego prawdopodobieństwa p^* .

Dziękuję za uwagę