

Binarne zmienne zależne

Zajęcia 15

Stanisław Cichocki

Natalia Nehrebecka

Plan zajęć

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

Plan zajęć

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

Binarne zmienne zależne

- zmienne zero-jedynkowe nazywane są także **zmiennymi binarnymi**
- dla wielu interesujących modeli ekonomicznych zmienna zależna jest zmienną zero-jedynkową
- zmienna zero-jedynkowa opisuje stan w jakim znajdują się badane obiekty

Binarne zmienne zależne

- Przykład:

W badaniach na danych przekrojowych dotyczących aktywności zawodowej ludności może nas interesować jakie są przyczyny bezrobocia.

Zmienną objaśnianą w tym przypadku będzie status zawodowy respondenta opisany zmienną zero-jedynkową (0 nie pracuje, 1 pracuje).

Binarne zmienne zależne

- przy kodowaniu zmiennej przyjmuje się konwencję:

- 0 - porażka

- 1 - sukces

- czym jest sukces a czym porażka zależy od badania

Binarne zmienne zależne

- Przykład:

0 - brak pracy, porażka

1 - wykonywanie pracy, sukces

Binarne zmienne zależne

- w modelach dla zmiennych binarnych wyjaśniamy, za pomocą zmiennych niezależnych, prawdopodobieństwa stanów
- by sformułować model musimy opisać prawdopodobieństwa zdarzeń dla poszczególnych obserwacji
- na największym poziomie ogólności:

$$\Pr(y_i | \mathbf{x}_i) = \begin{cases} 1 - p(\mathbf{x}_i) & \text{dla } y_i = 0 \\ p(\mathbf{x}_i) & \text{dla } y_i = 1 \end{cases}$$

gdzie $p(\mathbf{x}_i)$ opisuje zależność prawdopodobieństwa sukcesu od wielkości zmiennych niezależnych $\mathbf{x}_i$

Binarne zmienne zależne

- modele dla zmiennych binarnych nie są modelami liniowymi

Plan zajęć

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

Liniowy model prawdopodobieństwa (LPM)

- rozpatrujemy różne formy funkcyjne dla zależności między prawdopodobieństwem sukcesu a wielkością zmiennych objaśniających
- najprostszy model $\longrightarrow$ liniowy model prawdopodobieństwa (LPM)
- $p(x_i) = F(x_i\beta) = x_i\beta$

gdzie:

$p(x_i)$ - prawdopodobieństwo sukcesu $F()$ - dystrybuanta

Liniowy model prawdopodobieństwa (LPM)

- liniowy w tym sensie, że warunkowa wartość oczekiwana jest dana funkcją liniową:

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

Liniowy model prawdopodobieństwa (LPM)

- zalety LPM:

- a) łatwy do wyestymowania
- b) współczynniki = efekty cząstkowe

- wady LPM:

- a) wartości dopasowane spoza $[0;1]$ $\longrightarrow$ nieinterpretowalne
(prawdopodobieństwo sukcesu)
- b) heteroscedastyczność

Liniowy model prawdopodobieństwa (LPM)

- problemu z punktu a) nie da się w prosty sposób rozwiązać
- w przypadku heteroscedastyczności stosujemy SUMNK lub odporne macierze wariancji i kowariancji

Plan zajęć

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

LPM - Interpretacja współczynników

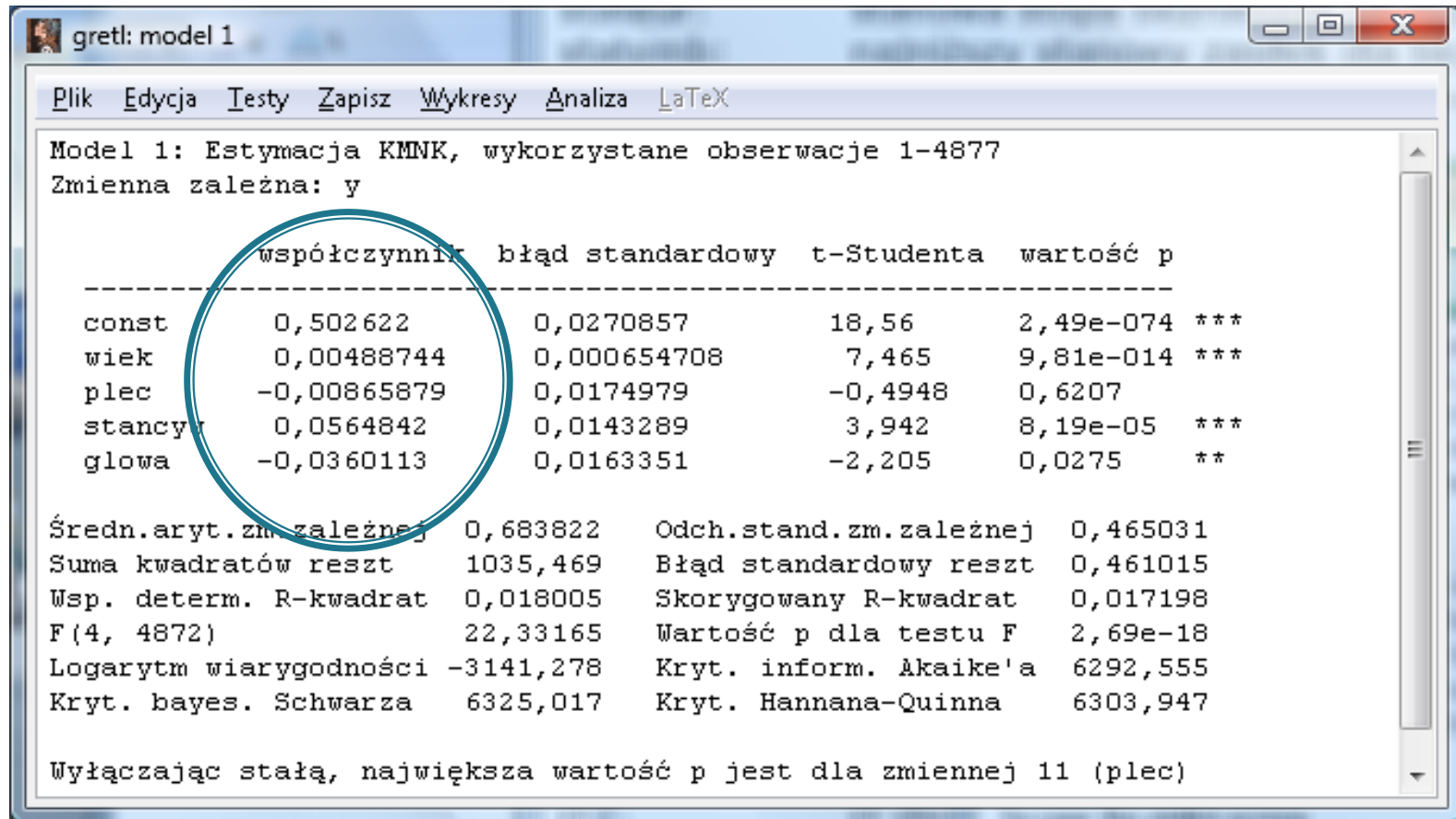
- w liniowym modelu prawdopodobieństwa interpretuje się współczynniki
- interpretacja:
 - a) **dla zmiennych objaśniających ciągłych:** wpływ jednostkowej zmiany zmiennej niezależnej na wielkość prawdopodobieństwa sukcesu
 - b) **dla zmiennych objaśniających zero-jedynkowych:**
różnica między prawdopodobieństwem sukcesu dla zmiennej zero-jedynkowej równej 0 i równej 1, przy pozostałych zmiennych ustalonych na poziomie średnich

LPM - Interpretacja współczynników

- Przykład:

- ▶ Modelowano odpowiedź na pytanie:
 - czy ubiegał się o zasiłek?
- ▶ Zmienna zależna y :
 - 0 – nie ubiegał się, 1 – ubiegał się.
- ▶ Zmienne niezależne:
 - wiek,
 - płeć (0-kobieta, 1 – mężczyzna),
 - stan cywilny (0 - pozostałe stany, 1 – zamężna/zamężny),
 - głowa (0 – nie jest głową rodziny, 1 – jest głową rodziny)

LPM - Interpretacja współczynników



gretl: model 1

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Model 1: Estymacja KLMN, wykorzystane obserwacje 1-4877
Zmienna zależna: y

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	0,502622	0,0270857	18,56	2,49e-074	***
wiek	0,00488744	0,000654708	7,465	9,81e-014	***
plec	-0,00865879	0,0174979	-0,4948	0,6207	
stancya	0,0564842	0,0143289	3,942	8,19e-05	***
glowa	-0,0360113	0,0163351	-2,205	0,0275	**

Średn. aryt. zm. zależnej	0,683822	Odch. stand. zm. zależnej	0,465031
Suma kwadratów reszt	1035,469	Błąd standardowy reszt	0,461015
Wsp. determ. R-kwadrat	0,018005	Skorygowany R-kwadrat	0,017198
F(4, 4872)	22,33165	Wartość p dla testu F	2,69e-18
Logarytm wiarygodności	-3141,278	Kryt. inform. Akaike'a	6292,555
Kryt. bayes. Schwarza	6325,017	Kryt. Hannana-Quinna	6303,947

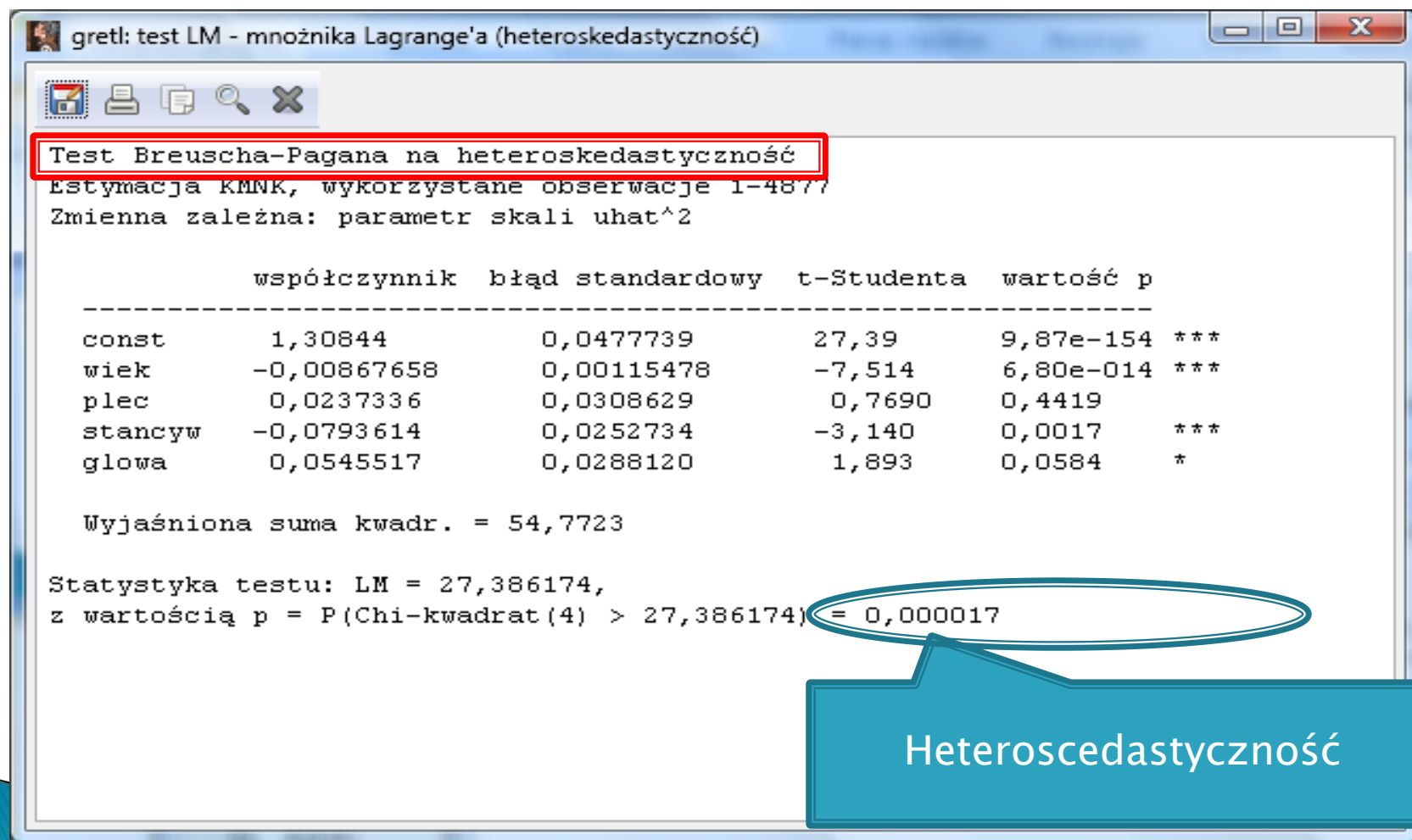
Wyłączając stałą, największa wartość p jest dla zmiennej 11 (plec)

LPM - Interpretacja współczynników

- interpretacja współczynników:

- a) wiek: wzrost wieku o rok powoduje wzrost prawdopodobieństwa ubiegania się o zasiłek o 0,4 punktu procentowego
- b) płeć: mężczyźni mają o 0,8 punktu procentowego mniejsze prawdopodobieństwo ubiegania się o zasiłek niż kobiety
- c) stan cywilny: osoby zamężne mają o 5,6 punktów procentowych większe prawdopodobieństwo ubiegania się o zasiłek niż pozostałe osoby
- d) głowa: osoba będąca głową rodziny ma o 3,6 punktów procentowych mniejsze prawdopodobieństwo ubiegania się o zasiłek niż osoby nie będące głowami rodzin

LPM – Problem heteroscedastyczności



Plan zajęć

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

Probit

- w modelu **probitowym** zakładamy, że $F(\cdot)$ jest dystrybucją rozkładu normalnego

Plan zajęć

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

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
 - 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

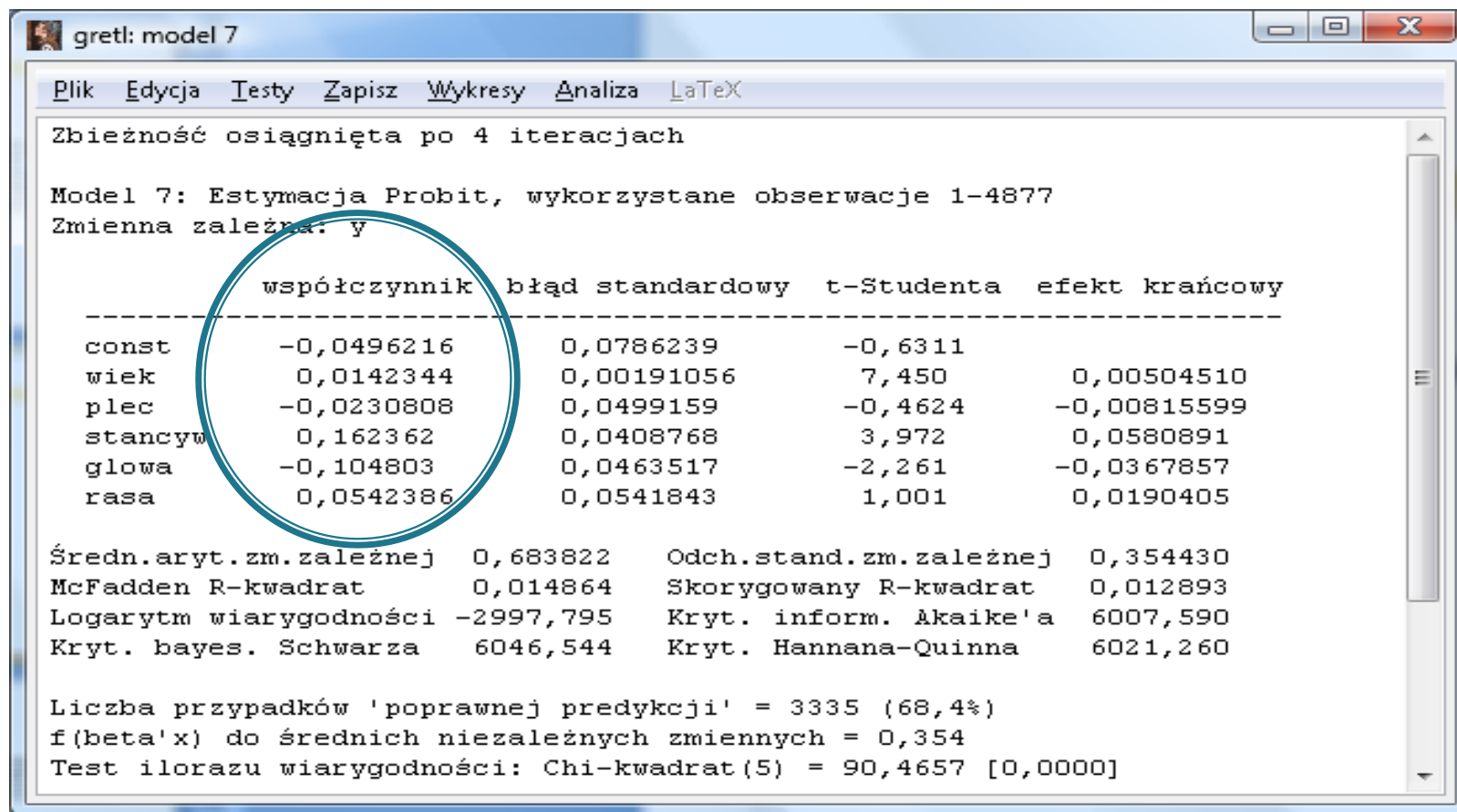
- efekty częściowe dla zmiennych objaśniających ciągłych liczymy zwykle dla średnich wartości tych zmiennych (efekty częściowe zależą od wielkości zmiennych objaśniających)
- znak efektu częściowego dla danej zmiennej jest taki sam jak znak współczynnika przy tej zmiennej
- możemy zatem interpretować znaki przy współczynnikach:
 - dodatni znak $\longrightarrow$ zmienna wpływa dodatnio na prawdopodobieństwo sukcesu
 - ujemny znak $\longrightarrow$ zmienna wpływa ujemnie na prawdopodobieństwo sukcesu

Probit - Interpretacja współczynników

- Przykład:

- ▶ Modelowano odpowiedź na pytanie:
 - czy ubiegał się o zasiłek?
- ▶ Zmienna zależna y :
 - 0 – nie ubiegał się, 1 – ubiegał się.
- ▶ Zmienne niezależne:
 - wiek,
 - płeć (0-kobieta, 1 – mężczyzna),
 - stan cywilny (0 - pozostałe stany, 1 – zamężna/zamężny),
 - głowa (0 – nie jest głową rodziny, 1 – jest głową rodziny)

Probit - Interpretacja współczynników



gretl: model 7

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Zbieżność osiągnięta po 4 iteracjach

Model 7: Estymacja Probit, wykorzystane obserwacje 1-4877
Zmienna zależna: y

	współczynnik	błąd standardowy	t-Studenta	efekt krańcowy
const	-0,0496216	0,0786239	-0,6311	
wiek	0,0142344	0,00191056	7,450	0,00504510
plec	-0,0230808	0,0499159	-0,4624	-0,00815599
stancym	0,162362	0,0408768	3,972	0,0580891
glowa	-0,104803	0,0463517	-2,261	-0,0367857
rasa	0,0542386	0,0541843	1,001	0,0190405

Średn. aryt. zm. zależnej	0,683822	Odch. stand. zm. zależnej	0,354430
McFadden R-kwadrat	0,014864	Skorygowany R-kwadrat	0,012893
Logarytm wiarygodności	-2997,795	Kryt. inform. Akaike'a	6007,590
Kryt. bayes. Schwarza	6046,544	Kryt. Hannana-Quinna	6021,260

Liczba przypadków 'poprawnej predykcji' = 3335 (68,4%)
f(beta'x) do średnich niezależnych zmiennych = 0,354
Test ilorazu wiarygodności: Chi-kwadrat(5) = 90,4657 [0,0000]

Probit - Interpretacja znaków wsp.

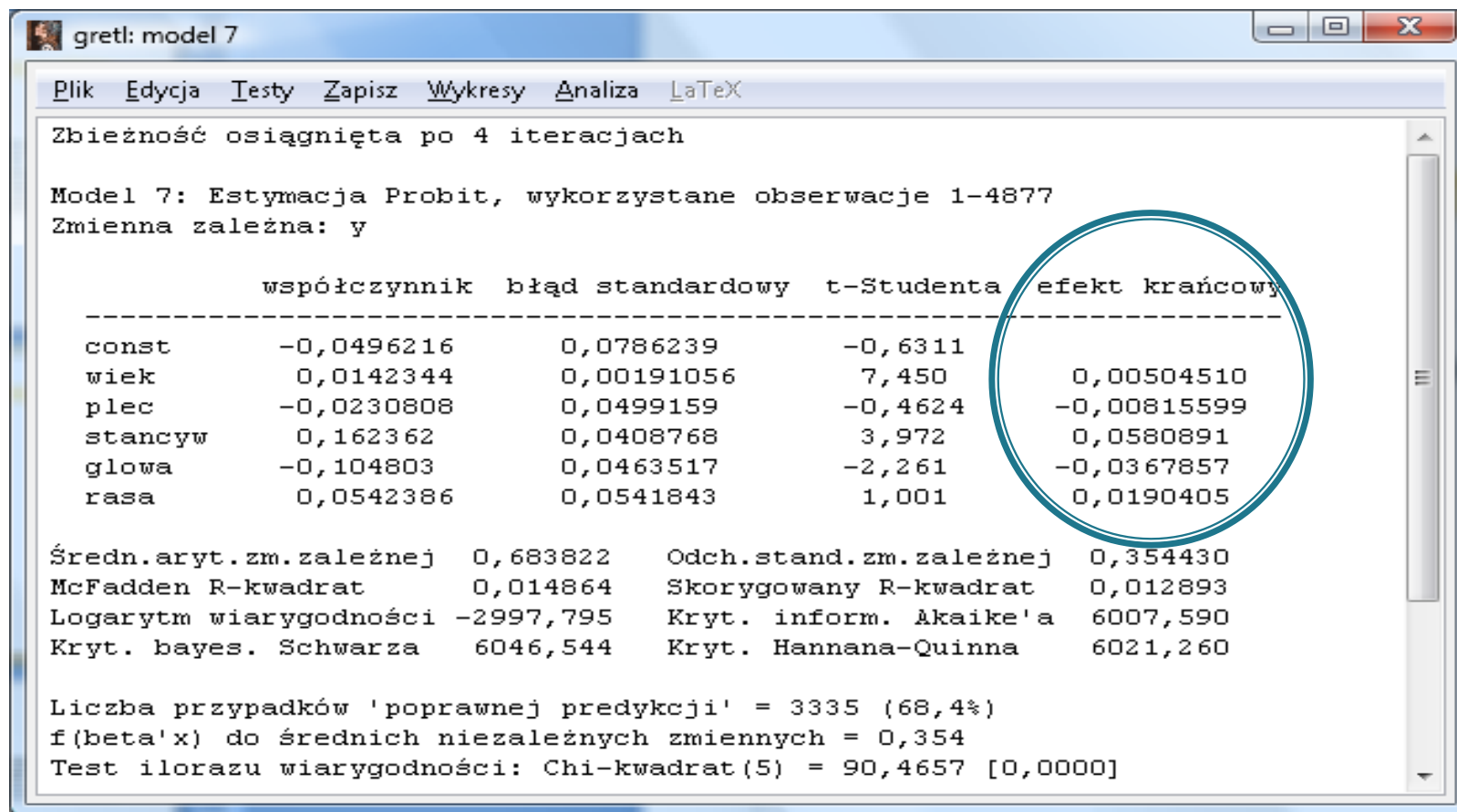
- interpretacja znaków:

- a) wiek: wpływa dodatnio na prawdopodobieństwo ubiegania się o zasiłek, im większy wiek tym większe prawdopodobieństwo ubiegania się o zasiłek
- b) płeć: wpływa ujemnie na prawdopodobieństwo ubiegania się o zasiłek, mężczyźni mają mniejsze prawdopodobieństwo ubiegania się o zasiłek niż kobiety
- c) stan cywilny: wpływa dodatnio na prawdopodobieństwo ubiegania się o zasiłek, osoby zamężne mają większe prawdopodobieństwo ubiegania się o zasiłek niż pozostałe osoby

Probit - Interpretacja znaków wsp.

- d) glowa: wpływa ujemnie na prawdopodobieństwo ubiegania się o zasiłek, osoba będąca głową rodziny ma mniejsze prawdopodobieństwo ubiegania się o zasiłek niż osoby nie będące głowami rodzin
- e) rasa: wpływa dodatnio na prawdopodobieństwo ubiegania się o zasiłek, osoby innych ras mają większe prawdopodobieństwo ubiegania się o zasiłek niż osoby rasy białej

Probit – Interpretacja efektów cząstkowych (krańcowych)



gretl: model 7

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Zbieżność osiągnięta po 4 iteracjach

Model 7: Estymacja Probit, wykorzystane obserwacje 1-4877
Zmienna zależna: y

	współczynnik	błąd standardowy	t-Studenta	efekt krańcowy
const	-0,0496216	0,0786239	-0,6311	
wiek	0,0142344	0,00191056	7,450	0,00504510
plec	-0,0230808	0,0499159	-0,4624	-0,00815599
stancyw	0,162362	0,0408768	3,972	0,0580891
glowa	-0,104803	0,0463517	-2,261	-0,0367857
rasa	0,0542386	0,0541843	1,001	0,0190405

Średn. aryt. zm. zależnej 0,683822 Odch. stand. zm. zależnej 0,354430
McFadden R-kwadrat 0,014864 Skorygowany R-kwadrat 0,012893
Logarytm wiarygodności -2997,795 Kryt. inform. Akaike'a 6007,590
Kryt. bayes. Schwarza 6046,544 Kryt. Hannana-Quinna 6021,260

Liczba przypadków 'poprawnej predykcji' = 3335 (68,4%)
f(beta'x) do średnich niezależnych zmiennych = 0,354
Test ilorazu wiarygodności: Chi-kwadrat(5) = 90,4657 [0,0000]

Probit – Interpretacja efektów cząstkowych (krańcowych)

- Interpretacja efektów cząstkowych (krańcowych)

a) wiek: wzrost wieku o rok powoduje wzrost prawdopodobieństwa ubiegania się o zasiłek o 0,5 punktu procentowego

b) płeć: mężczyźni mają o 0,8 punktu procentowego mniejsze prawdopodobieństwo ubiegania się o zasiłek niż kobiety

c) stan cywilny: osoby zamężne mają o 5,8 punktów procentowych większe prawdopodobieństwo ubiegania się o zasiłek niż pozostałe osoby

d) głowa: osoba będąca głową rodziny ma o 3,6 punktów procentowych mniejsze prawdopodobieństwo ubiegania się o zasiłek niż osoby nie będące głowami rodzin

Probit – Interpretacja efektów cząstkowych (krańcowych)

e) rasa: osoba rasy innej niż biała ma o 1,9 punktów procentowych większe prawdopodobieństwo ubiegania się o zasiłek niż osoby rasy białej

Probit – test na łączną istotność zmiennych w modelu

gretl: model 7

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Zbieżność osiągnięta po 4 iteracjach

Model 7: Estymacja Probit, wykorzystane obserwacje 1-4877
Zmienna zależna: y

	współczynnik	błąd standardowy	t-Studenta	efekt krańcowy
const	-0,0496216	0,0786239	-0,6311	
wiek	0,0142344	0,00191056	7,450	0,00504510
plec	-0,0230808	0,0499159	-0,4624	-0,00815599
stancyw	0,162362	0,0408768	3,972	0,0580891
glowa	-0,104803	0,0463517	-2,261	-0,0367857
rasa	0,0542386	0,0541843	1,001	0,0190405

Średn. aryt. zm. zależnej	0,683822	Odch. stand. zm. zależnej	0,354430
Mcfadden R-kwadrat	0,014864	Skorygowany R-kwadrat	0,012893
Logarytm wiarygodności	-2997,795	Kryt. inform. Akaike'a	6007,590
Kryt. bayes. Schwarza	6046,544	Kryt. Hannana-Quinna	6021,260

Liczba przypadków 'poprawnej predykcji' = 3335 (68,4%)
f(beta'x) do średnich niezależnych zmiennych = 0,354
Test ilorazu wiarygodności: Chi-kwadrat(5) = 90,4657 [0,0000]

Łączna istotność zmiennych

Plan zajęć

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

Probit - Miary dopasowania

- tablica klasyfikacyjna:

- analiza częstości wartości y_i^* zaobserwowanych i przewidzianych

- zakładamy, że model przewiduje sukces, jeśli dopasowane z modelu prawdopodobieństwo sukcesu y_i jest większe o progowego prawdopodobieństwa p^*

Probit - Miary dopasowania

- tablica klasyfikacyjna:

		Przewidywane (Predicted)	
		0	1
Empiryczne (Actual)	0	N 00	N 01
	1	N 10	N 11

Prognostujemy porażkę i
obserwujemy sukces

Prognostujemy porażkę i
obserwujemy porażkę

Probit - Miary dopasowania

- na podstawie tablicy klasyfikacyjnej można ustalić na ile model trafnie przewiduje sukcesy i porażki
- liczebnościowe R^2 : **proporcja zdarzeń (sukcesów bądź porażek) prawidłowo przewidywanych przez model**

$$R^2_{\text{liczebnościowe}} = \frac{N00 + N11}{N00 + N01 + N10 + N11}$$

Probit - 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 :

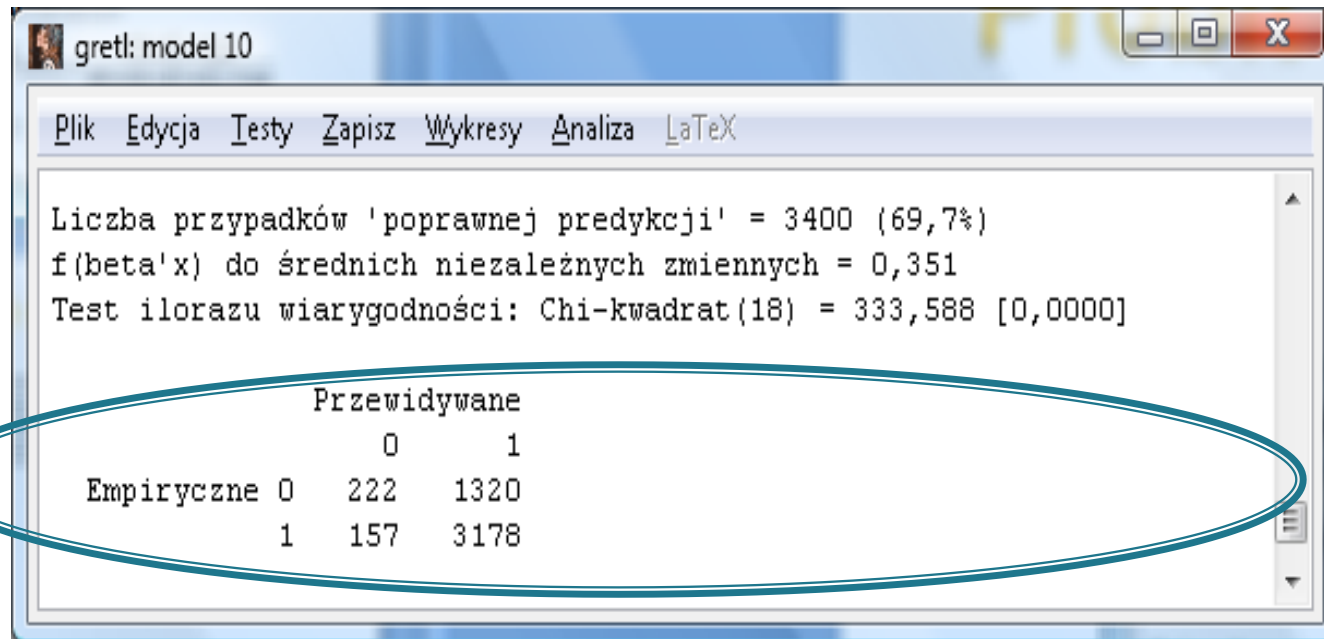
$$\overline{R^2}_{\text{liczebnościowe}} = \frac{N_{00} + N_{11} - N_{\max}}{N_{00} + N_{01} + N_{10} + N_{11} - N_{\max}}$$

- Gdzie $N_{\max}$ - liczba obserwacji dla zdarzenia
(sukces/porażka), którą obserwowano częściej

Probit - Miary dopasowania

- skorygowane liczebnościowe R^2 $\longrightarrow$ 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

Probit – Miary dopasowania



gretl: model 10

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Liczba przypadków 'poprawnej predykcji' = 3400 (69,7%)
f(beta'x) do średnich niezależnych zmiennych = 0,351
Test ilorazu wiarygodności: Chi-kwadrat(18) = 333,588 [0,0000]

		Przewidywane	
		0	1
Empiryczne	0	222	1320
	1	157	3178

Probit - Miary dopasowania

- liczebnościowe R^2 :

$$R^2_{\text{liczebnościowe}} = \frac{222 + 3178}{222 + 1320 + 157 + 3178} \approx 0,697$$

- skorygowane liczebnościowe R^2 :

$$\bar{R}^2 = \frac{222 + 3178 - (3178 + 157)}{222 + 1320 + 157 + 3178 - (3178 + 157)} = 0,035$$

Probit - Miary dopasowania

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

$$wrazliwosc = \frac{N11}{N10 + N11}$$

Probit - Miary dopasowania

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

$$specyficznosc = \frac{N00}{N00 + N01}$$

Probit - Miary dopasowania

- Przykład:

Możliwe są dwa przypadki prawidłowych wskazań w przypadku testu na obecność wirusa: pozytywny wynik testu w przypadku osoby, która ma wirusa i negatywny w przypadku osoby, w której organizmie nie ma wirusa. Wrażliwość testu na obecność wirusa HIV jest więc udziałem osób, dla których test dał wynik pozytywny wśród osób, które mają wirusa, a specyficzność testu będzie udziałem osób, dla których test dał wynik negatywny wśród tych osób, które nie mają wirusa.

Probit - Miary dopasowania

- dodatkowo:

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

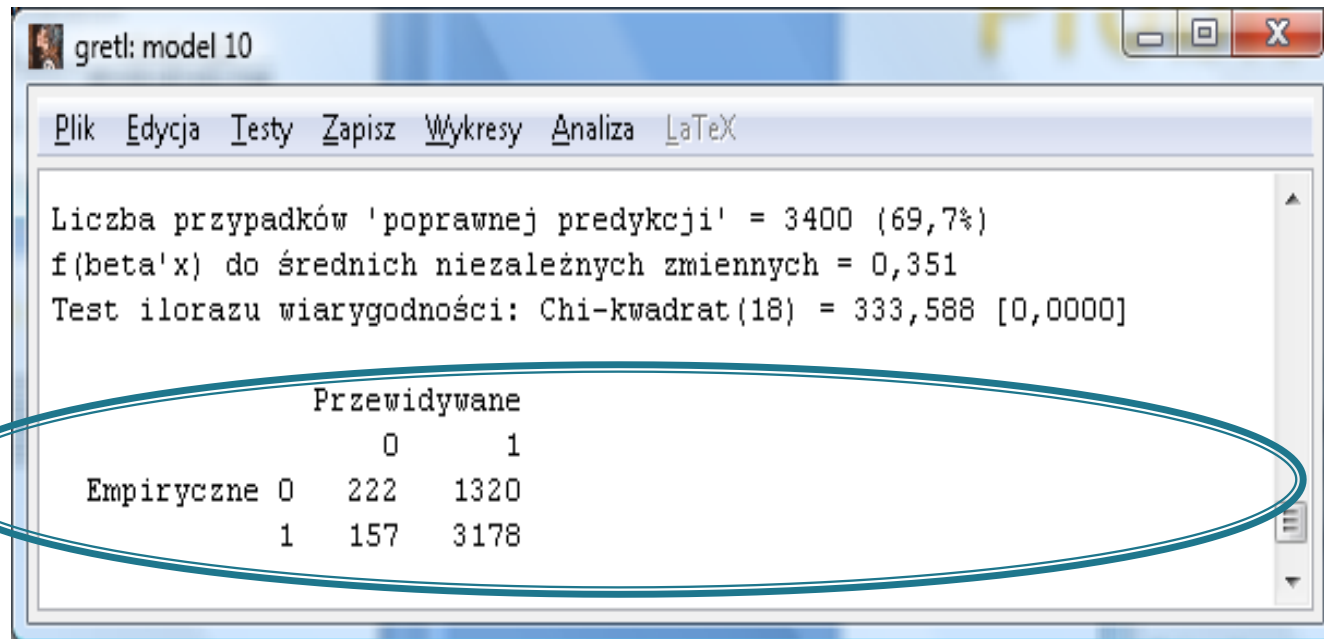
$$1 - \text{wrazliwosc} = \frac{N10}{N10 + N11}$$

Probit - Miary dopasowania

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

$$1 - \textit{specyficzność} = \frac{N01}{N00 + N01}$$

Probit – Miary dopasowania



gretl: model 10

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Liczba przypadków 'poprawnej predykcji' = 3400 (69,7%)
f(beta'x) do średnich niezależnych zmiennych = 0,351
Test ilorazu wiarygodności: Chi-kwadrat(18) = 333,588 [0,0000]

		Przewidywane	
		0	1
Empiryczne	0	222	1320
	1	157	3178

Probit - Miary dopasowania

$$wrazliwosc = \frac{3178}{157 + 3178} \approx 0,95$$

$$specyficznosc = \frac{222}{222 + 1320} \approx 0,14$$

$$1 - wrazliwosc = \frac{157}{157 + 3178} \approx 0,04$$

$$1 - specyficznosc = \frac{1320}{222 + 1320} \approx 0,85$$

Plan zajęć

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

Logit

- w **modelu logitowym** zakładamy, że $F()$ jest dystrybuantą rozkładu logistycznego

- $$p(x_i) = F(x_i\beta) = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}}$$

Plan zajęć

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

Logit - 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
 - 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**

Logit - Interpretacja współczynników

- efekty częściowe dla zmiennych objaśniających ciągłych liczymy zwykle dla średnich wartości tych zmiennych (efekty częściowe zależą od wielkości zmiennych objaśniających)
- znak efektu częściowego dla danej zmiennej jest taki sam jak znak współczynnika przy tej zmiennej
- możemy zatem interpretować znaki przy współczynnikach:
 - dodatni znak ➡ zmienna wpływa dodatnio na prawdopodobieństwo sukcesu
 - ujemny znak ➡ zmienna wpływa ujemnie na prawdopodobieństwo sukcesu

Logit - Interpretacja współczynników

- Przykład:

- ▶ Modelowano odpowiedź na pytanie:
 - czy ubiegał się o zasiłek?
- ▶ Zmienna zależna y :
 - 0 – nie ubiegał się, 1 – ubiegał się.
- ▶ Zmienne niezależne:
 - wiek,
 - płeć (0-kobieta, 1 – mężczyzna),
 - stan cywilny (0 - pozostałe stany, 1 – zamężna/zamężny),
 - głowa (0 – nie jest głową rodziny, 1 – jest głową rodziny)

Logit - Interpretacja współczynników

gretl: model 13

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Zbieżność osiągnięta po 4 iteracjach

Model 13: Estymacja Logit, wykorzystane obserwacje 1-4877

Zmienna zależna: y

	współczynnik	błąd standardowy	t-Studenta	efekt krańcowy
const	-0,0190718	0,136777	-0,1394	
wiek	0,0239035	0,00321020	7,446	0,00513448
plec	-0,0325247	0,0822594	-0,3954	-0,00696362
stancjw	0,259055	0,0673776	3,845	0,0563055
glowa	-0,181917	0,0763397	-2,383	-0,0385739
rasa	0,0918313	0,0895727	1,025	0,0194808
miasto	-0,129239	0,0661415	-1,954	-0,0275481

Średn. aryt. zm. zależnej 0,683822 Odch. stand. zm. zależnej 0,214800

McFadden R-kwadrat 0,015528 Skorygowany R-kwadrat 0,013228

Logarytm wiarygodności -2995,776 Kryt. inform. Akaike'a 6005,551

Kryt. bayes. Schwarza 6050,997 Kryt. Hannana-Quinna 6021,500

Liczba przypadków 'poprawnej predykcji' = 3335 (68,4%)

f(beta'x) do średnich niezależnych zmiennych = 0,215

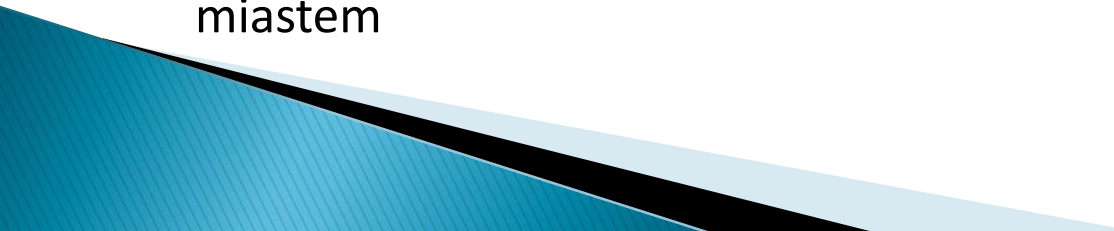
Test ilorazu wiarygodności: Chi-kwadrat(6) = 94,5046 [0,0000]

Logit - Interpretacja znaków wsp.

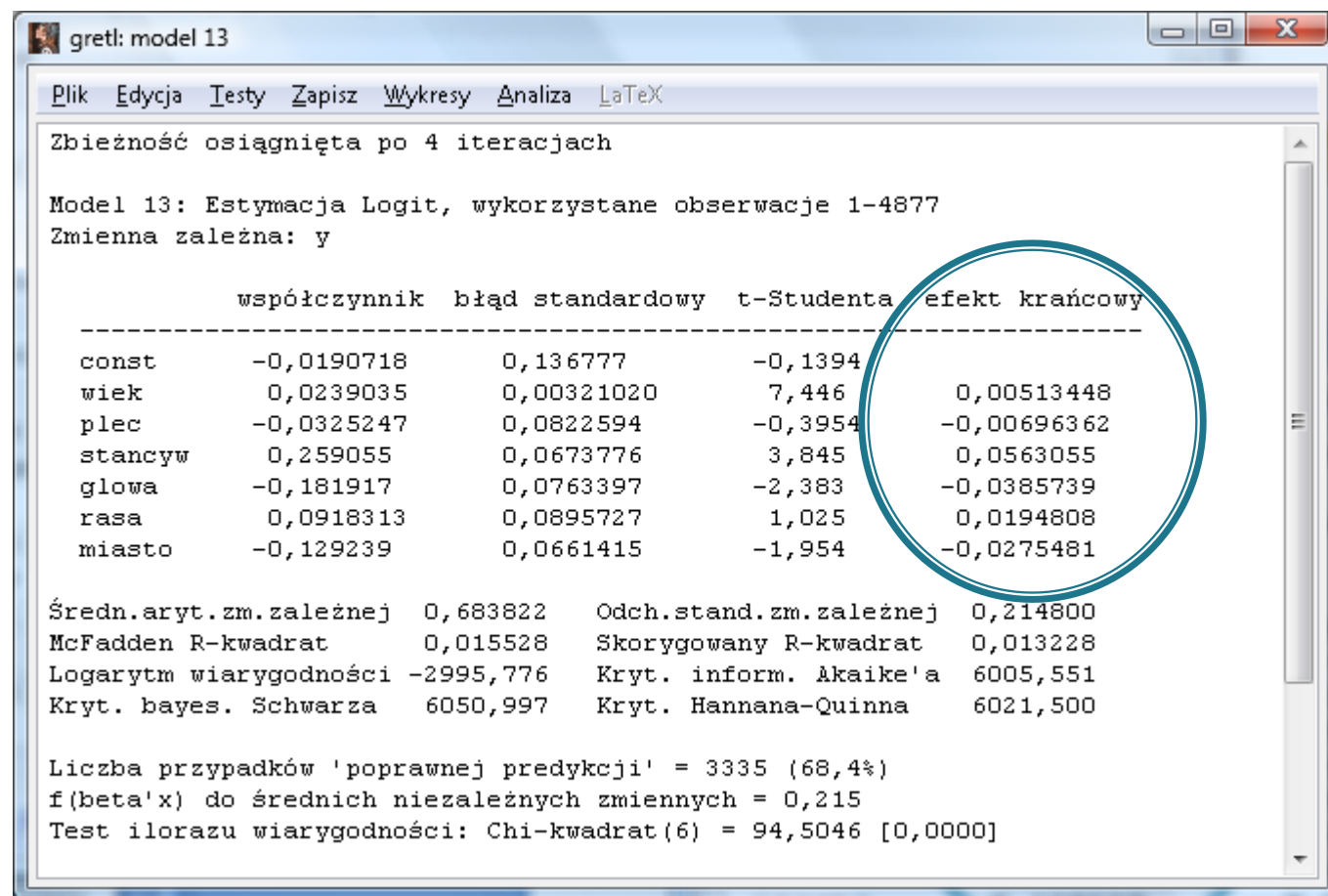
- interpretacja znaków:

- a) wiek: wpływa dodatnio na prawdopodobieństwo ubiegania się o zasiłek, im większy wiek tym większe prawdopodobieństwo ubiegania się o zasiłek
- b) płeć: wpływa ujemnie na prawdopodobieństwo ubiegania się o zasiłek, mężczyźni mają mniejsze prawdopodobieństwo ubiegania się o zasiłek niż kobiety
- c) stan cywilny: wpływa dodatnio na prawdopodobieństwo ubiegania się o zasiłek, osoby zamężne mają większe prawdopodobieństwo ubiegania się o zasiłek niż pozostałe osoby

Logit - Interpretacja znaków wsp.

- d) glowa: wpływa ujemnie na prawdopodobieństwo ubiegania się o zasiłek, osoba będąca głową rodziny ma mniejsze prawdopodobieństwo ubiegania się o zasiłek niż osoby nie będące głowami rodzin
 - e) rasa: wpływa dodatnio na prawdopodobieństwo ubiegania się o zasiłek, osoby innych ras mają większe prawdopodobieństwo ubiegania się o zasiłek niż osoby rasy białej
 - f) miasto: wpływa ujemnie na prawdopodobieństwo ubiegania się o zasiłek, osoby mieszkające w miastach mają mniejsze prawdopodobieństwo ubiegania się o zasiłek niż mieszkające poza miastem
- 

Logit – Interpretacja efektów częstokowych (krańcowych)



gretl: model 13

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Zbieżność osiągnięta po 4 iteracjach

Model 13: Estymacja Logit, wykorzystane obserwacje 1-4877
Zmienna zależna: y

	współczynnik	błąd standardowy	t-Studenta	efekt krańcowy
const	-0,0190718	0,136777	-0,1394	
wiek	0,0239035	0,00321020	7,446	0,00513448
plec	-0,0325247	0,0822594	-0,3954	-0,00696362
stancyw	0,259055	0,0673776	3,845	0,0563055
glowa	-0,181917	0,0763397	-2,383	-0,0385739
rasa	0,0918313	0,0895727	1,025	0,0194808
miasto	-0,129239	0,0661415	-1,954	-0,0275481

Średn. aryt. zm. zależnej	0,683822	Odch. stand. zm. zależnej	0,214800
McFadden R-kwadrat	0,015528	Skorygowany R-kwadrat	0,013228
Logarytm wiarygodności	-2995,776	Kryt. inform. Akaike'a	6005,551
Kryt. bayes. Schwarza	6050,997	Kryt. Hannana-Quinna	6021,500

Liczba przypadków 'poprawnej predykcji' = 3335 (68,4%)
f(beta'x) do średnich niezależnych zmiennych = 0,215
Test ilorazu wiarygodności: Chi-kwadrat(6) = 94,5046 [0,0000]

Logit – Interpretacja efektów cząstkowych (krańcowych)

- Interpretacja efektów cząstkowych (krańcowych)

a) wiek: wzrost wieku o rok powoduje wzrost prawdopodobieństwa ubiegania się o zasiłek o 0,5 punktu procentowego

b) płeć: mężczyźni mają o 0,6 punktu procentowego mniejsze prawdopodobieństwo ubiegania się o zasiłek niż kobiety

c) stan cywilny: osoby zamężne mają o 5,6 punktów procentowych większe prawdopodobieństwo ubiegania się o zasiłek niż pozostałe osoby

d) głowa: osoba będąca głową rodziny ma o 3,8 punktów procentowych mniejsze prawdopodobieństwo ubiegania się o zasiłek niż osoby nie będące głowami rodzin

Logit – Interpretacja efektów cząstkowych (krańcowych)

- e) rasa: osoba rasy innej niż biała ma o 1,9 punktów procentowych większe prawdopodobieństwo ubiegania się o zasiłek niż osoby rasy białej
- e) miasto: osoba mieszkająca w mieście ma o 2,7 punktów procentowych mniejsze prawdopodobieństwo ubiegania się o zasiłek niż osoba mieszkająca poza miastem

Probit – Interpretacja efektów cząstkowych (krańcowych)

gretl: model 13

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Zbieżność osiągnięta po 4 iteracjach

Model 13: Estymacja Logit, wykorzystane obserwacje 1-4877
Zmienna zależna: y

	współczynnik	błąd standardowy	t-Studenta	efekt krańcowy
const	-0,0190718	0,136777	-0,1394	
wiek	0,0239035	0,00321020	7,446	0,00513448
plec	-0,0325247	0,0822594	-0,3954	-0,00696362
stancyw	0,259055	0,0673776	3,845	0,0563055
glowa	-0,181917	0,0763397	-2,383	-0,0385739
rasa	0,0918313	0,0895727	1,025	0,0194808
miasto	-0,129239	0,0661415	-1,954	-0,0275481

Średn. aryt. zm. zależnej	0,683822	Odch. stand. zm. zależnej	0,214800
Mcfadden R-kwadrat	0,015528	Skorygowany R-kwadrat	0,013228
Logarytm wiarygodności	-2995,776	Kryt. inform. Akaike'a	6005,551
Kryt. bayes. Schwarza	6050,997	Kryt. Hannana-Quinna	6021,500

Liczba przypadków 'poprawnej predykcji' = 3335 (68,4%)
~~f(beta'y) do średnich niezależnych zmiennych = 0,215~~
Test ilorazu wiarygodności: Chi-kwadrat(6) = 94,5046 [0,0000]

Łączna istotność zmiennych

Plan zajęć

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

Logit -Miary dopasowania

- takie same jak w modelu probitowym

Pytania teoretyczne

1. Co jest modelowane w przypadku jeśli zmienna zależna jest zmienna binarną (zero-jedynkową)?
2. Jakie są wady liniowego modelu prawdopodobieństwa?
3. Wyjaśnić czym jest wrażliwość i specyficzność.

Dziękuję za uwagę