

## Dyskretne zmienne zależne

- Zmienna dyskretna to zmienna przyjmująca skończoną liczbę ("małą") wartości

- Zmienna dyskretna to zmienna przyjmująca skończoną liczbę ("małą") wartości
- W praktyce oznacza się je za pomocą kolejnych liczb naturalnych

- Zmienna dyskretna to zmienna przyjmująca skończoną liczbę ("małą") wartości
- W praktyce oznacza się je za pomocą kolejnych liczb naturalnych
- Celem budowy modelu jest oszacowanie warunkowego prawdopodobieństwa wystąpienia jednego ze stanów

- Przyjmijmy, że modelowana zmienna zależna przyjmuje jedną z  $J$  wartości

- Przyjmijmy, że modelowana zmienna zależna przyjmuje jedną z  $J$  wartości
- Każda z wartości opisuje jeden stan zjawiska

- Przyjmijmy, że modelowana zmienna zależna przyjmuje jedną z  $J$  wartości
- Każda z wartości opisuje jeden stan zjawiska
- Na przykład, gdy modelujemy stan cywilny osoby

1	kawaler, panna
2	żonaty, mężatka
3	wdowiec, wdowa
4	rozwódziona(a)

- W zależności od charakteru zjawiska może wystąpić jedna z dwóch sytuacji

- W zależności od charakteru zjawiska może wystąpić jedna z dwóch sytuacji
- Wartości zmiennej zależnej można w naturalny sposób uporządkować

- W zależności od charakteru zjawiska może wystąpić jedna z dwóch sytuacji
- Wartości zmiennej zależnej można w naturalny sposób uporządkować
- Zmienna nie posiada narzuconego porządku

- Zakładamy, że obserwowana zmienna dyskretna jest determinowana przez nieobserwowany proces (indeks)

$$y_i^* = X_i\beta + \varepsilon_i$$

gdzie  $\varepsilon_i$  ma rozkład  $F(\cdot)$

- Zakładamy, że obserwowana zmienna dyskretna jest determinowana przez nieobserwowany proces (indeks)

$$y_i^* = X_i\beta + \varepsilon_i$$

gdzie  $\varepsilon_i$  ma rozkład  $F(\cdot)$

- W celu identyfikacji parametrów modelu przyjmuje się, że  $var(\varepsilon_i) = 1$

- Zakładamy, że obserwowana zmienna dyskretna jest determinowana przez nieobserwowany proces (indeks)

$$y_i^* = X_i\beta + \varepsilon_i$$

gdzie  $\varepsilon_i$  ma rozkład  $F(\cdot)$

- W celu identyfikacji parametrów modelu przyjmuje się, że  $\text{var}(\varepsilon_i) = 1$
- Związek między obserwowaną zmienną zależną  $y_i$  a zmienną ukrytą  $y_i^*$

$y_i = 0$	$y_i^* \leq \alpha_1$
$y_i = 1$	$\alpha_1 \leq y_i^* \leq \alpha_2$
$\vdots$	$\vdots$
$y_i = J$	$\alpha_J \leq y_i^*$

- Zakładamy, że obserwowana zmienna dyskretna jest determinowana przez nieobserwowany proces (indeks)

$$y_i^* = X_i\beta + \varepsilon_i$$

gdzie  $\varepsilon_i$  ma rozkład  $F(\cdot)$

- W celu identyfikacji parametrów modelu przyjmuje się, że  $\text{var}(\varepsilon_i) = 1$
- Związek między obserwowaną zmienną zależną  $y_i$  a zmienną ukrytą  $y_i^*$

$y_i = 0$	$y_i^* \leq \alpha_1$
$y_i = 1$	$\alpha_1 \leq y_i^* \leq \alpha_2$
$\vdots$	$\vdots$
$y_i = J$	$\alpha_J \leq y_i^*$

- Nieznane parametry  $\alpha_j$  nazywamy punktami odcięcia.

- Punkty odcięcia muszą spełniać nierówność

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_J$$

- Punkty odcięcia muszą spełniać nierówność

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_J$$

- Prawdopodobieństwo wyboru alternatywy  $j$  wynosi

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} < y_i^* \leq \alpha_j)$$

- Punkty odcięcia muszą spełniać nierówność

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_J$$

- Prawdopodobieństwo wyboru alternatywy  $j$  wynosi

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} < y_i^* \leq \alpha_j)$$

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} < X_i\beta + \varepsilon_i \leq \alpha_j)$$

- Punkty odcięcia muszą spełniać nierówność

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_J$$

- Prawdopodobieństwo wyboru alternatywy  $j$  wynosi

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} < y_i^* \leq \alpha_j)$$

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} < X_i\beta + \varepsilon_i \leq \alpha_j)$$

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} - X_i\beta < \varepsilon_i \leq \alpha_j - X_i\beta)$$

- Punkty odcięcia muszą spełniać nierówność

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_J$$

- Prawdopodobieństwo wyboru alternatywy  $j$  wynosi

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} < y_i^* \leq \alpha_j)$$

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} < X_i\beta + \varepsilon_i \leq \alpha_j)$$

$$Pr(y_i = j|X) = Pr(\alpha_{j-1} - X_i\beta < \varepsilon_i \leq \alpha_j - X_i\beta)$$

$$Pr(y_i = j|X) = F(\alpha_j - X_i\beta) - F(\alpha_{j-1} - X_i\beta)$$

- Prawdopodobieństwa wyboru poszczególnych alternatyw wynoszą odpowiednio

$$Pr(y_i = 0|X) = Pr(y_i^* < \alpha_1|X) = F(\alpha_1 - X\beta)$$

- Prawdopodobieństwa wyboru poszczególnych alternatyw wynoszą odpowiednio

$$Pr(y_i = 0|X) = Pr(y_i^* < \alpha_1|X) = F(\alpha_1 - X\beta)$$

$$Pr(y_i = 1|X) = Pr(\alpha_1 \leq y_i^* < \alpha_2|X) = F(\alpha_2 - X\beta) - F(\alpha_1 - X\beta)$$

⋮

$$Pr(y_i = J|X) = Pr(\alpha_J \leq y_i^*) = 1 - F(\alpha_J - X\beta)$$

- Prawdopodobieństwa wyboru poszczególnych alternatyw są niezależne i muszą sumować się do 1.

- Prawdopodobieństwa wyboru poszczególnych alternatyw są niezależne i muszą sumować się do 1.
- Postać funkcji wiarygodności zależy od postaci dystrybuanty  $F(\cdot)$

- Prawdopodobieństwa wyboru poszczególnych alternatyw są niezależne i muszą sumować się do 1.
- Postać funkcji wiarygodności zależy od postaci dystrybuanty  $F(\cdot)$
- Zazwyczaj przyjmuje się, że  $F(\cdot)$  ma rozkład normalny lub logistyczny

- Prawdopodobieństwa wyboru poszczególnych alternatyw są niezależne i muszą sumować się do 1.
- Postać funkcji wiarygodności zależy od postaci dystrybuanty  $F(\cdot)$
- Zazwyczaj przyjmuje się, że  $F(\cdot)$  ma rozkład normalny lub logistyczny
- W przypadku rozkładu normalnego model będziemy nazywać uporządkowanym probitem

- Prawdopodobieństwa wyboru poszczególnych alternatywy są niezależne i muszą sumować się do 1.
- Postać funkcji wiarygodności zależy od postaci dystrybuanty  $F(\cdot)$
- Zazwyczaj przyjmuje się, że  $F(\cdot)$  ma rozkład normalny lub logistyczny
- W przypadku rozkładu normalnego model będziemy nazywać uporządkowanym probitem
- W przypadku rozkładu logistycznego model będziemy nazywać uporządkowanym logitem

- W modelach z dyskretną zmienną zależną efekty cząstkowe są definiowane oddzielnie dla każdej alternatywy

$$\frac{\partial Pr(y_i = 1)}{\partial x} = -\beta_0 f(\alpha_0 - X_i \beta)$$

- W modelach z dyskretną zmienną zależną efekty cząstkowe są definiowane oddzielnie dla każdej alternatywy

$$\frac{\partial Pr(y_i = 1)}{\partial x} = -\beta_0 f(\alpha_0 - X_i \beta)$$

$$\frac{\partial Pr(y_i = j)}{\partial x} = \beta_j [f(\alpha_{j-1} - X_i \beta) - f(\alpha_j - X_i \beta)]$$

- W modelach z dyskretną zmienną zależną efekty częściowe są definiowane oddzielnie dla każdej alternatywy

$$\frac{\partial Pr(y_i = 1)}{\partial x} = -\beta_0 f(\alpha_0 - X_i \beta)$$

$$\frac{\partial Pr(y_i = j)}{\partial x} = \beta_j [f(\alpha_{j-1} - X_i \beta) - f(\alpha_j - X_i \beta)]$$

$$\frac{\partial Pr(y_i = J)}{\partial x} = \beta_J f(\alpha_J - X_i \beta)$$

- W modelach z dyskretną zmienną zależną efekty częściowe są definiowane oddzielnie dla każdej alternatywy

$$\frac{\partial \Pr(y_i = 1)}{\partial x} = -\beta_0 f(\alpha_0 - X_i \beta)$$

$$\frac{\partial \Pr(y_i = j)}{\partial x} = \beta_j [f(\alpha_{j-1} - X_i \beta) - f(\alpha_j - X_i \beta)]$$

$$\frac{\partial \Pr(y_i = J)}{\partial x} = \beta_J f(\alpha_J - X_i \beta)$$

- Znaki efektów częściowych związanych z alternatywami skrajnymi są determinowane przez znaki parametrów

- W modelach z dyskretną zmienną zależną efekty cząstkowe są definiowane oddzielnie dla każdej alternatywy

$$\frac{\partial Pr(y_i = 1)}{\partial x} = -\beta_0 f(\alpha_0 - X_i \beta)$$

$$\frac{\partial Pr(y_i = j)}{\partial x} = \beta_j [f(\alpha_{j-1} - X_i \beta) - f(\alpha_j - X_i \beta)]$$

$$\frac{\partial Pr(y_i = J)}{\partial x} = \beta_J f(\alpha_J - X_i \beta)$$

- Znaki efektów cząstkowych związanych z alternatywami skrajnymi są determinowane przez znaki parametrów
- Dla pośrednich alternatyw nie ma takiego związku

- Model opisuje prawdopodobieństwo wyboru "wyższych" kategorii w stosunku do niższych

- Model opisuje prawdopodobieństwo wyboru "wyższych" kategorii w stosunku do niższych
- Model uporządkowany jest modelem jednorównaniowym

- Model opisuje prawdopodobieństwo wyboru "wyższych" kategorii w stosunku do niższych
- Model uporządkowany jest modelem jednorównaniowym
- Zatem wpływ zmiennej niezależnej jest opisywany przed jeden współczynnik

- Model opisuje prawdopodobieństwo wyboru "wyższych" kategorii w stosunku do niższych
- Model uporządkowany jest modelem jednorównaniowym
- Zatem wpływ zmiennej niezależnej jest opisywany przed jeden współczynnik
- Aby mieć możliwość interpretacji zakłada się, że zachodzi tzw. równoległa regresja

- Model opisuje prawdopodobieństwo wyboru "wyższych" kategorii w stosunku do niższych
- Model uporządkowany jest modelem jednorównaniowym
- Zatem wpływ zmiennej niezależnej jest opisywany przed jeden współczynnik
- Aby mieć możliwość interpretacji zakłada się, że zachodzi tzw. równoległa regresja

$$Pr(y_i \leq 1|X) = F(\alpha_1 - \beta_0 - \sum_k X_k \beta_k)$$

- Model opisuje prawdopodobieństwo wyboru "wyższych" kategorii w stosunku do niższych
- Model uporządkowany jest modelem jednorównaniowym
- Zatem wpływ zmiennej niezależnej jest opisywany przed jeden współczynnik
- Aby mieć możliwość interpretacji zakłada się, że zachodzi tzw. równoległa regresja

$$Pr(y_i \leq 1|X) = F(\alpha_1 - \beta_0 - \sum_k X_k \beta_k)$$

$$Pr(y_i \leq 2|X) = F(\alpha_2 - \beta_0 - \sum_k X_k \beta_k)$$

- Model opisuje prawdopodobieństwo wyboru "wyższych" kategorii w stosunku do niższych
- Model uporządkowany jest modelem jednorównaniowym
- Zatem wpływ zmiennej niezależnej jest opisywany przed jeden współczynnik
- Aby mieć możliwość interpretacji zakłada się, że zachodzi tzw. równoległa regresja

$$Pr(y_i \leq 1|X) = F(\alpha_1 - \beta_0 - \sum_k X_k \beta_k)$$

$$Pr(y_i \leq 2|X) = F(\alpha_2 - \beta_0 - \sum_k X_k \beta_k)$$

- Więc alternatywy są rozróżniane poprzez wartość stałej

- Jest to poważne ograniczenie modelu uporządkowanego

- Jest to poważne ograniczenie modelu uporządkowanego
- Niestety nie istnieje formalna powszechnie akceptowana procedura testowa

- Jest to poważne ograniczenie modelu uporządkowanego
- Niestety nie istnieje formalna powszechnie akceptowana procedura testowa
- S.Long opracował test brandt

- Jest to poważne ograniczenie modelu uporządkowanego
- Niestety nie istnieje formalna powszechnie akceptowana procedura testowa
- S.Long opracował test brandt
- Porównuje oszacowane wartości współczynników modelu uporządkowanego z oszacowaniami współczynników odpowiednich modeli z binarną zmienną zależną

# Oszacowania parametrów

Ordered logistic regression

Number of obs = 6812  
 LR chi2(6) = 472.46  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.0527

Log likelihood = -4248.9393

stcyw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
wiek	.0777431	.0049097	15.83	0.000	.0681203	.087366
_Iwyksz_2	.2974954	.0981448	3.03	0.002	.1051352	.4898556
_Iwyksz_3	.273511	.1651773	1.66	0.098	-.0502306	.5972527
_Iwyksz_4	-.0757362	.0877578	-0.86	0.388	-.2477384	.096266
_Iwyksz_5	-1.109434	.1126925	-9.84	0.000	-1.330308	-.8885612
wies	-.2435972	.0614689	-3.96	0.000	-.3640741	-.1231203
/cut1	1.393604	.1963353			1.008793	1.778414
/cut2	6.321276	.221512			5.887121	6.755432
/cut3	6.458236	.2227267			6.021699	6.894772

# Oszacowania ilorazów szans

Ordered logistic regression

Number of obs = 6812  
 LR chi2(6) = 472.46  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.0527

Log likelihood = -4248.9393

stcyw	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
wiek	1.080845	.0053066	15.83	0.000	1.070494	1.091296
_Iwyksz_2	1.346482	.1321502	3.03	0.002	1.110861	1.632081
_Iwyksz_3	1.314572	.2171375	1.66	0.098	.9510101	1.81712
_Iwyksz_4	.9270608	.0813569	-0.86	0.388	.7805641	1.101052
_Iwyksz_5	.3297454	.0371598	-9.84	0.000	.2643959	.411247
wies	.7838033	.0481796	-3.96	0.000	.6948397	.8841573
/cut1	1.393604	.1963353			1.008793	1.778414
/cut2	6.321276	.221512			5.887121	6.755432
/cut3	6.458236	.2227267			6.021699	6.894772

Variable	ologit	oprobit
-----+-----		
stcyw		
wiek	.07774314	.03999982
_Iwyksz_2	.29749543	.15509029
_Iwyksz_3	.27351103	.15948366
_Iwyksz_4	-.07573617	-.03198528
_Iwyksz_5	-1.1094344	-.55379973
wies	-.24359717	-.14474071
-----+-----		
cut1		
_cons	1.3936036	.59224007
-----+-----		
cut2		
_cons	6.3212765	3.3448399
-----+-----		
cut3		
_cons	6.4582358	3.4046196
-----		

# Test Branta

## Brant Test of Parallel Regression Assumption

Variable	chi2	p>chi2	df
All	158.74	0.000	12
wiek	19.42	0.000	2
_Iwyksz_2	3.79	0.150	2
_Iwyksz_3	8.37	0.015	2
_Iwyksz_4	21.50	0.000	2
_Iwyksz_5	75.94	0.000	2
wies	6.70	0.035	2

A significant test statistic provides evidence that the parallel regression assumption has been violated.

## Test formy funkcyjnej (linktest)

Ordered logistic regression

Number of obs = 6812  
 LR chi2(2) = 474.59  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.0529

Log likelihood = -4247.871

stcyw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_hat	1.419664	.2917318	4.87	0.000	.8478802	1.991448
_hatsq	-.077267	.0529371	-1.46	0.144	-.1810218	.0264878
/cut1	1.931856	.3931408			1.161315	2.702398
/cut2	6.84612	.3968768			6.068256	7.623985
/cut3	6.983048	.3975291			6.203905	7.76219

- Dyskretna zmienna zależna nie zawsze jest zmienną porządkową

- Dyskretna zmienna zależna nie zawsze jest zmienną porządkową
- W takim przypadku będziemy stosować model wielomianowy

- Dyskretna zmienna zależna nie zawsze jest zmienną porządkową
- W takim przypadku będziemy stosować model wielomianowy
- Analogicznie jak w przypadku modelu uporządkowanego wybór dystrybucyjny sprowadza się do wyboru pomiędzy rozkładem normalnym i logistycznym

- Dyskretna zmienna zależna nie zawsze jest zmienną porządkową
- W takim przypadku będziemy stosować model wielomianowy
- Analogicznie jak w przypadku modelu uporządkowanego wybór dystrybucyjny sprowadza się do wyboru pomiędzy rozkładem normalnym i logistycznym
- Z uwagi na łatwość obliczeń i interpretacji rozkład logistyczny jest preferowany

- Załóżmy, że zmienna zależna przyjmuje jedną z wartości  $\{0, \dots, J\}$

- Załóżmy, że zmienna zależna przyjmuje jedną z wartości  $\{0, \dots, J\}$
- Prawdopodobieństwo wyboru konkretnej alternatywy jest dane przez

$$Pr(y = j|X) = \begin{cases} \frac{1}{1 + \sum_{i=1}^J \exp(X_i\beta)} & j = 0 \\ \frac{\exp(X_j\beta)}{1 + \sum_{i=1}^J \exp(X_i\beta)} & j = 1 \dots J \end{cases}$$

- Załóżmy, że zmienna zależna przyjmuje jedną z wartości  $\{0, \dots, J\}$
- Prawdopodobieństwo wyboru konkretnej alternatywy jest dane przez

$$Pr(y = j|X) = \begin{cases} \frac{1}{1 + \sum_{i=1}^J \exp(X_i\beta)} & j = 0 \\ \frac{\exp(X_i\beta)}{1 + \sum_{i=1}^J \exp(X_i\beta)} & j = 1 \dots J \end{cases}$$

- Poziom  $y_0$  zmiennej zależnej jest interpretowany jako poziom bazowy zjawiska

- Oszacowania parametrów i efektów cząstkowych są interpretowane w odniesieniu do poziomu bazowego zmiennej zależnej

- Oszacowania parametrów i efektów cząstkowych są interpretowane w odniesieniu do poziomu bazowego zmiennej zależnej
- Można również parametry modelu przedstawić w postaci ilorazów szans

- Oszacowania parametrów i efektów cząstkowych są interpretowane w odniesieniu do poziomu bazowego zmiennej zależnej
- Można również parametry modelu przedstawić w postaci ilorazów szans
- Szansa jest równa stosunkowi prawdopodobieństwa zaobserwowania  $J$  alternatywy do prawdopodobieństwa wystąpienia poziomu bazowego

$$Odds(x) = \frac{Pr(y = J)}{Pr(y = 0)} = \exp(X_i\beta)$$

- Iloraz szans jest dany wzorem

$$\frac{Odds_i(x)}{Odds_j(x)} = \frac{\frac{Pr(y=i)}{Pr(y=0)}}{\frac{Pr(y=j)}{Pr(y=0)}} = \frac{Pr(y=i)}{Pr(y=j)}$$

- Iloraz szans jest dany wzorem

$$\frac{Odds_i(x)}{Odds_j(x)} = \frac{\frac{Pr(y=i)}{Pr(y=0)}}{\frac{Pr(y=j)}{Pr(y=0)}} = \frac{Pr(y=i)}{Pr(y=j)}$$

$$\frac{Odds_i(x)}{Odds_j(x)} = \exp(X\beta_i - X\beta_j) = \exp(x\Delta\beta)$$

- Iloraz szans jest dany wzorem

$$\frac{Odds_i(x)}{Odds_j(x)} = \frac{\frac{Pr(y=i)}{Pr(y=0)}}{\frac{Pr(y=j)}{Pr(y=0)}} = \frac{Pr(y=i)}{Pr(y=j)}$$

$$\frac{Odds_i(x)}{Odds_j(x)} = \exp(X\beta_i - X\beta_j) = \exp(x\Delta\beta)$$

- Dodatnia wartość ilorazu szans wskazuje, że wraz ze wzrostem wartości cechy  $x$  rośnie relatywne prawdopodobieństwo zaobserwowania wyniku  $y_i$  względem wyniku  $y_j$

- Dla każdej alternatywy można obliczyć efekty cząstkowe mierzące wpływ poszczególnych zmiennych niezależnych na prawdopodobieństwo zaobserwowania alternatywy

$$\frac{\partial Pr(y = j)}{\partial x} = \frac{\beta_j \exp(x_j \beta_j) [1 + \sum \exp(x \beta_h)] - \exp(x_j \beta_j) [\sum \beta_h \exp(x \beta_h)]}{[1 + \sum \exp(X \beta_h)]^2}$$

- Dla każdej alternatywy można obliczyć efekty cząstkowe mierzące wpływ poszczególnych zmiennych niezależnych na prawdopodobieństwo zaobserwowania alternatywy

$$\frac{\partial Pr(y = j)}{\partial x} = \frac{\beta_j \exp(x_j \beta_j) [1 + \sum \exp(x \beta_h)] - \exp(x_j \beta_j) [\sum \beta_h \exp(x \beta_h)]}{[1 + \sum \exp(X \beta_h)]^2}$$

- Zatem znaki efektów cząstkowych zależą od wartości  $x$  dla których zostały obliczone

- Dla każdej alternatywy można obliczyć efekty cząstkowe mierzące wpływ poszczególnych zmiennych niezależnych na prawdopodobieństwo zaobserwowania alternatywy

$$\frac{\partial Pr(y = j)}{\partial x} = \frac{\beta_j \exp(x_j \beta_j) [1 + \sum \exp(x \beta_h)] - \exp(x_j \beta_j) [\sum \beta_h \exp(x \beta_h)]}{[1 + \sum \exp(X \beta_h)]^2}$$

- Zatem znaki efektów cząstkowych zależą od wartości  $x$  dla których zostały obliczone
- Jest to powód dla którego z reguły analizuje się ilorazy szans

## Wyniki (1/4)

Multinomial logistic regression

Log likelihood = -4192.9097

Number of obs	=	6812
LR chi2(18)	=	584.52
Prob > chi2	=	0.0000
Pseudo R2	=	0.0652

## Wyniki (2/4)

	stcyw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1							
	wiek	.918038	.0050073	-15.68	0.000	.908276	.9279048
	_Iwyksz_2	.7249511	.0828227	-2.82	0.005	.5795119	.9068908
	_Iwyksz_3	.9504316	.1716669	-0.28	0.778	.6670784	1.354144
	_Iwyksz_4	1.248905	.1220669	2.27	0.023	1.031178	1.512603
	_Iwyksz_5	4.15078	.4965289	11.90	0.000	3.283269	5.247507
	wies	1.172674	.0795568	2.35	0.019	1.026668	1.339445

(stcyw==2 is the base outcome)

## Wyniki (3/4)

3							
wiek		1.086522	.0360318	2.50	0.012	1.018147	1.159488
_Iwyksz_2		2.49e+08	3.78e+08	12.75	0.000	1.28e+07	4.87e+09
_Iwyksz_3		6.55e+08	1.02e+09	13.07	0.000	3.12e+07	1.38e+10
_Iwyksz_4		4.95e+08	7.41e+08	13.37	0.000	2.63e+07	9.32e+09
_Iwyksz_5		1.50e+09	2.30e+09	13.80	0.000	7.47e+07	3.02e+10
wies		.5490114	.200156	-1.64	0.100	.2686917	1.121782

(stcyw==2 is the base outcome)

## Wyniki (4/4)

---

4							
wiek		1.02016	.0113711	1.79	0.073	.9981144	1.042692
_Iwyksz_2		1.35217	.3213795	1.27	0.204	.8486338	2.154478
_Iwyksz_3		2.276279	.7344278	2.55	0.011	1.209457	4.28411
_Iwyksz_4		1.491598	.3307489	1.80	0.071	.9658401	2.303554
_Iwyksz_5		2.892317	.8022866	3.83	0.000	1.679326	4.981461
wies		.5980091	.0860658	-3.57	0.000	.4510269	.7928904

---

(stcyw==2 is the base outcome)

- Powodem uwzględnienia poziomu bazowego w modelu wielomianowym jest identyfikacja parametrów modelu

- Powodem uwzględnienia poziomu bazowego w modelu wielomianowym jest identyfikacja parametrów modelu
- W modelu nie można zidentyfikować bezpośrednio prawdopodobieństw, a jedynie wpływ każdej zmiennej objaśniającej różnicę w częstości występowania dwóch wyników

- Powodem uwzględnienia poziomu bazowego w modelu wielomianowym jest identyfikacja parametrów modelu
- W modelu nie można zidentyfikować bezpośrednio prawdopodobieństw, a jedynie wpływ każdej zmiennej objaśniającej różnicę w częstości występowania dwóch wyników
- Dowód

$$Pr(y = j|X) = \frac{\exp(x\beta_j)}{\sum_h \exp(x\beta_h)} = \frac{\frac{\exp(x\beta_j)}{\exp(x\beta_0)}}{\frac{\sum_h \exp(x\beta_h)}{\exp(x\beta_0)}}$$

- Powodem uwzględnienia poziomu bazowego w modelu wielomianowym jest identyfikacja parametrów modelu
- W modelu nie można zidentyfikować bezpośrednio prawdopodobieństw, a jedynie wpływ każdej zmiennej objaśniającej różnicę w częstości występowania dwóch wyników
- Dowód

$$Pr(y = j|X) = \frac{\exp(x\beta_j)}{\sum_h \exp(x\beta_h)} = \frac{\frac{\exp(x\beta_j)}{\exp(x\beta_0)}}{\frac{\sum_h \exp(x\beta_h)}{\exp(x\beta_0)}}$$

$$Pr(y = j|X) = \frac{\exp(x(\beta_j - \beta))}{\sum_h \exp(x(\beta_h - \beta))} = Pr(y = j|X)$$

- więc wektory  $\beta_j$  oraz  $\beta_j - \beta$  generują takie same prawdopodobieństwa dla alternatyw

- więc wektory  $\beta_j$  oraz  $\beta_j - \beta$  generują takie same prawdopodobieństwa dla alternatyw
- Dla uzyskania jednoznacznych rozwiązań przyjmuje się  $\beta_0 = 0$

- więc wektory  $\beta_j$  oraz  $\beta_j - \beta$  generują takie same prawdopodobieństwa dla alternatyw
- Dla uzyskania jednoznacznych rozwiązań przyjmuje się  $\beta_0 = 0$
- Wtedy parametry pokazują różnicę między daną alternatywą a poziomem bazowym zjawiska

- Ze wzoru na iloraz szans wynika, że relatywne prawdopodobieństwo wystąpienia alternatywy  $j$  nie zależy od prawdopodobieństw wystąpienia pozostałych alternatyw

- Ze wzoru na iloraz szans wynika, że relatywne prawdopodobieństwo wystąpienia alternatywy  $j$  nie zależy od prawdopodobieństw wystąpienia pozostałych alternatyw
- To założenie jest nazywane niezależnością niezwiązanych alternatyw

- Ze wzoru na iloraz szans wynika, że relatywne prawdopodobieństwo wystąpienia alternatywy  $j$  nie zależy od prawdopodobieństw wystąpienia pozostałych alternatyw
- To założenie jest nazywane niezależnością niezwiązanych alternatyw
- Jest to największa słabość modelu wielomianowego i warunkowego (o modelu warunkowym nie będziemy mówić)

- Ze wzoru na iloraz szans wynika, że relatywne prawdopodobieństwo wystąpienia alternatywy  $j$  nie zależy od prawdopodobieństw wystąpienia pozostałych alternatyw
- To założenie jest nazywane niezależnością niezwiązanych alternatyw
- Jest to największa słabość modelu wielomianowego i warunkowego (o modelu warunkowym nie będziemy mówić)
- Wynika z niego, że dodanie kolejnej alternatywy nie wpływa na relatywne prawdopodobieństwo alternatyw wcześniej uwzględnionych

## Przykład - McFadden

- Przypuśćmy, że z miasta A do miasta B możemy odbyć podróż samochodem lub autobusem czerwonej linii

## Przykład - McFadden

- Przypuśćmy, że z miasta A do miasta B możemy odbyć podróż samochodem lub autobusem czerwonej linii
- Załóżmy, że pasażerowie z równym prawdopodobieństwem wybierają środek transportu

## Przykład - McFadden

- Przypuśćmy, że z miasta A do miasta B możemy odbyć podróż samochodem lub autobusem czerwonej linii
- Załóżmy, że pasażerowie z równym prawdopodobieństwem wybierają środek transportu
- Więc  $Pr(\text{samochod}) = \frac{1}{2}$ ,  $Pr(\text{autobus}) = \frac{1}{2}$ ,  
 $Odds(s/a) = \frac{1/2}{1/2} = 1$

## Przykład - McFadden

- Przypuśćmy, że z miasta A do miasta B możemy odbyć podróż samochodem lub autobusem czerwonej linii
- Załóżmy, że pasażerowie z równym prawdopodobieństwem wybierają środek transportu
- Więc  $Pr(samochod) = \frac{1}{2}$ ,  $Pr(autobus) = \frac{1}{2}$ ,  
 $Odds(s/a) = \frac{1/2}{1/2} = 1$
- Przypuśćmy, że uruchomiono nową niebieską linię autobusową

## Przykład - McFadden

- Przypuśćmy, że z miasta A do miasta B możemy odbyć podróż samochodem lub autobusem czerwonej linii
- Załóżmy, że pasażerowie z równym prawdopodobieństwem wybierają środek transportu
- Więc  $Pr(samochod) = \frac{1}{2}$ ,  $Pr(autobus) = \frac{1}{2}$ ,  
 $Odds(s/a) = \frac{1/2}{1/2} = 1$
- Przypuśćmy, że uruchomiono nową niebieską linię autobusową
- Rozsądnie jest założyć, że po pojawieniu się nowego autobusu pasażerowie jeżdżący samochodem dalej będą to czynić

## Przykład - McFadden

- Przypuśćmy, że z miasta A do miasta B możemy odbyć podróż samochodem lub autobusem czerwonej linii
- Załóżmy, że pasażerowie z równym prawdopodobieństwem wybierają środek transportu
- Więc  $Pr(samochod) = \frac{1}{2}$ ,  $Pr(autobus) = \frac{1}{2}$ ,  
 $Odds(s/a) = \frac{1/2}{1/2} = 1$
- Przypuśćmy, że uruchomiono nową niebieską linię autobusową
- Rozsądnie jest założyć, że po pojawieniu się nowego autobusu pasażerowie jeżdżący samochodem dalej będą to czynić
- Więc  $Pr(samochod) = \frac{1}{2}$ ,  $Pr(autobus_c) = \frac{1}{4}$ ,  
 $Pr(autobus_n) = \frac{1}{4}$

## Przykład - McFadden

- Przypuśćmy, że z miasta A do miasta B możemy odbyć podróż samochodem lub autobusem czerwonej linii
- Załóżmy, że pasażerowie z równym prawdopodobieństwem wybierają środek transportu
- Więc  $Pr(samochod) = \frac{1}{2}$ ,  $Pr(autobus) = \frac{1}{2}$ ,  
 $Odds(s/a) = \frac{1/2}{1/2} = 1$
- Przypuśćmy, że uruchomiono nową niebieską linię autobusową
- Rozsądnie jest założyć, że po pojawieniu się nowego autobusu pasażerowie jeżdżący samochodem dalej będą to czynić
- Więc  $Pr(samochod) = \frac{1}{2}$ ,  $Pr(autobus_c) = \frac{1}{4}$ ,  
 $Pr(autobus_n) = \frac{1}{4}$
- Ale wówczas  $Odds(s/a_c) = \frac{1/2}{1/4} = 2$ ,  $Odds(s/a_n) = \frac{1/2}{1/4} = 2$

## Przykład - McFadden

- Oczywiście  $Odds(s/a) = \frac{1/2}{1/4+1/4} = 1$

## Przykład - McFadden

- Oczywiście  $Odds(s/a) = \frac{1/2}{1/4+1/4} = 1$
- Ale pojawienie się niebieskiej linii autobusowej zmieniło relatywne prawdopodobieństwo wyboru czerwonej linii autobusowej

## Przykład - McFadden

- Oczywiście  $Odds(s/a) = \frac{1/2}{1/4+1/4} = 1$
- Ale pojawienie się niebieskiej linii autobusowej zmieniło relatywne prawdopodobieństwo wyboru czerwonej linii autobusowej
- Jest to sprzeczne z założeniem o niezależności niezwiązanych alternatyw

## Przykład - McFadden

- Oczywiście  $Odds(s/a) = \frac{1/2}{1/4+1/4} = 1$
- Ale pojawienie się niebieskiej linii autobusowej zmieniło relatywne prawdopodobieństwo wyboru czerwonej linii autobusowej
- Jest to sprzeczne z założeniem o niezależności niezwiązanych alternatyw
- Aby założenie było spełnione, po pojawieniu się nowej linii autobusowej prawdopodobieństwa powinny wynosić odpowiednio  $Pr(samochod) = \frac{1}{3}$ ,  $Pr(autobus_c) = \frac{1}{3}$ ,  $Pr(autobus_n) = \frac{1}{3}$

- Ponieważ model jest wielorównaniowy oceniając istotność parametrów przy zmiennych należy badać łączną istotność parametrów we wszystkich równaniach
- Wystarczy, że zmienna będzie odróżniała 2 z  $J$  kategorii to będzie mogła być uznana za istotną
- Jako miar dopasowania używamy standardowych statystyk zaprojektowanych do modeli szacowanych MNW
- Należy pamiętać, że liczba szacowanych parametrów to liczba zmiennych mnożona przez liczbę równań

- Wartości dyskretnej zmiennej zależnej oznaczają natężenie zjawiska

- Wartości dyskretnej zmiennej zależnej oznaczają natężenie zjawiska
- Parametry takiego modelu posiadają interpretację ilościową

- Wartości dyskretnej zmiennej zależnej oznaczają natężenie zjawiska
- Parametry takiego modelu posiadają interpretację ilościową
- Do szacowania parametrów nie powinno stosować się KMRL, ponieważ wartości dopasowane mogą nie posiadać interpretacji

- Wartości dyskretnej zmiennej zależnej oznaczają natężenie zjawiska
- Parametry takiego modelu posiadają interpretację ilościową
- Do szacowania parametrów nie powinno stosować się KMRL, ponieważ wartości dopasowane mogą nie posiadać interpretacji
- Z tego powodu najczęściej wykorzystywanym modelem jest model Poissona

- Wartości dyskretnej zmiennej zależnej oznaczają natężenie zjawiska
- Parametry takiego modelu posiadają interpretację ilościową
- Do szacowania parametrów nie powinno stosować się KMRL, ponieważ wartości dopasowane mogą nie posiadać interpretacji
- Z tego powodu najczęściej wykorzystywanym modelem jest model Poissona
- Rozkład Poissona jest dany przez

$$Pr(y = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ dla } \lambda > 0$$

- W modelu Poissona

$$E(y) = \text{var}(y) = \lambda$$

- W modelu Poissona

$$E(y) = \text{var}(y) = \lambda$$

- Dane empiryczne rzadko spełniają powyższe założenie

- W modelu Poissona

$$E(y) = \text{var}(y) = \lambda$$

- Dane empiryczne rzadko spełniają powyższe założenie
- Bez utraty ogólności można założyć, że  $\lambda = \mu(x)$

- W modelu Poissona

$$E(y) = \text{var}(y) = \lambda$$

- Dane empiryczne rzadko spełniają powyższe założenie
- Bez utraty ogólności można założyć, że  $\lambda = \mu(x)$
- Zazwyczaj przyjmuje się, że  $\lambda = \exp(x)$ , z uwagi na wymóg dodatniej  $\lambda$

- Przy tych założeniach warunkowa wartość oczekiwana zmiennej zależnej wynosi

$$Pr(y = k|X) = \frac{\mu(x)^k}{k!} e^{-\mu(x)}$$

- Przy tych założeniach warunkowa wartość oczekiwana zmiennej zależnej wynosi

$$Pr(y = k|X) = \frac{\mu(x)^k}{k!} e^{-\mu(x)}$$

- Parametry modelu są semielastycznościami ponieważ

$$\frac{\partial \ln E(y|X)}{\partial x_k} = \frac{\partial \ln \mu(x)}{\partial x_k} = \beta_k$$

- Przy tych założeniach warunkowa wartość oczekiwana zmiennej zależnej wynosi

$$Pr(y = k|X) = \frac{\mu(x)^k}{k!} e^{-\mu(x)}$$

- Parametry modelu są semielastycznościami ponieważ

$$\frac{\partial \ln E(y|X)}{\partial x_k} = \frac{\partial \ln \mu(x)}{\partial x_k} = \beta_k$$

- Więc

$$\frac{\delta \mu(x)}{\mu(x)} / \Delta x_k \approx \beta_k$$

- Przy tych założeniach warunkowa wartość oczekiwana zmiennej zależnej wynosi

$$Pr(y = k|X) = \frac{\mu(x)^k}{k!} e^{-\mu(x)}$$

- Parametry modelu są semielastycznościami ponieważ

$$\frac{\partial \ln E(y|X)}{\partial x_k} = \frac{\partial \ln \mu(x)}{\partial x_k} = \beta_k$$

- Więc

$$\frac{\delta \mu(x)}{\mu(x)} / \Delta x_k \approx \beta_k$$

- Jeżeli  $x_k$  jest zlogarytmowane to parametr jest elastycznością

$$E(y) = \text{var}(y) = \lambda = \mu(x)$$

$$E(y) = \text{var}(y) = \lambda = \mu(x)$$

- Problemem jest spełnienie tego założenia przez dane empiryczne

$$E(y) = \text{var}(y) = \lambda = \mu(x)$$

- Problemem jest spełnienie tego założenia przez dane empiryczne
- Nawet gdy założenie jest niespełnione oszacowania są zgodne

$$E(y) = \text{var}(y) = \lambda = \mu(x)$$

- Problemem jest spełnienie tego założenia przez dane empiryczne
- Nawet gdy założenie jest niespełnione oszacowania są zgodne
- Przy niespełnionym założeniu do szacowania macierzy wariancji należy wykorzystać odporny estymator

Poisson regression

Number of obs = 53329328  
 LR chi2(6) = 2180682.72  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.0112

Log likelihood = -96476998

los	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Iregion_2	.0523322	.0002522	207.53	0.000	.051838	.0528265
_Iregion_3	.0486235	.000266	182.80	0.000	.0481022	.0491448
_Iregion_4	.0730661	.0002723	268.31	0.000	.0725323	.0735998
_Iregion_5	.010454	.0003104	33.68	0.000	.0098457	.0110623
_Iregion_6	.0447549	.0002787	160.59	0.000	.0442087	.0453012
wies	.2419851	.0001697	1426.15	0.000	.2416525	.2423177
_cons	.9152318	.000188	4866.96	0.000	.9148632	.9156003

```
. estat gof
```

```
Goodness-of-fit chi2 = 4.41e+07  
Prob > chi2(53329321) = 1.0000
```