Zastosowanie modelu potęgowego
Zmienne dyskretne

Stanisław Cichocki

Natalia Nehrebecka

Wykład 6

Plan wykładu

- 1. Zastosowanie modelu potęgowego
  - Przekształcenie Boxa-Coxa
- 2. Zmienne ciągłe za zmienne dyskretne
- 3. Interpretacja parametrów przy zmiennych dyskretnych

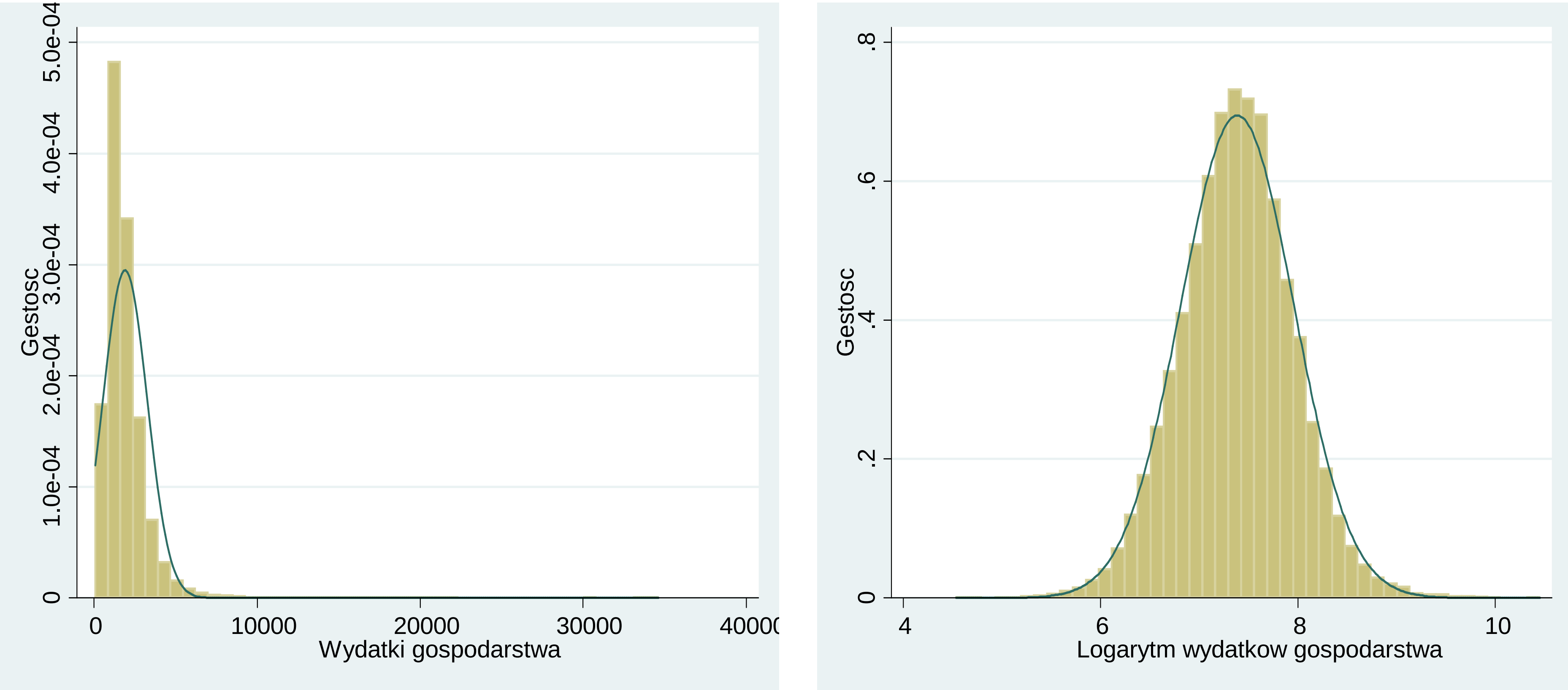
Plan wykładu

- 1. Zastosowanie modelu potęgowego
  - Przekształcenie Boxa-Coxa
- 2. Zmienne ciągłe za zmienne dyskretne
- 3. Interpretacja parametrów przy zmiennych dyskretnych

Zastosowanie modelu potęgowego - przykład

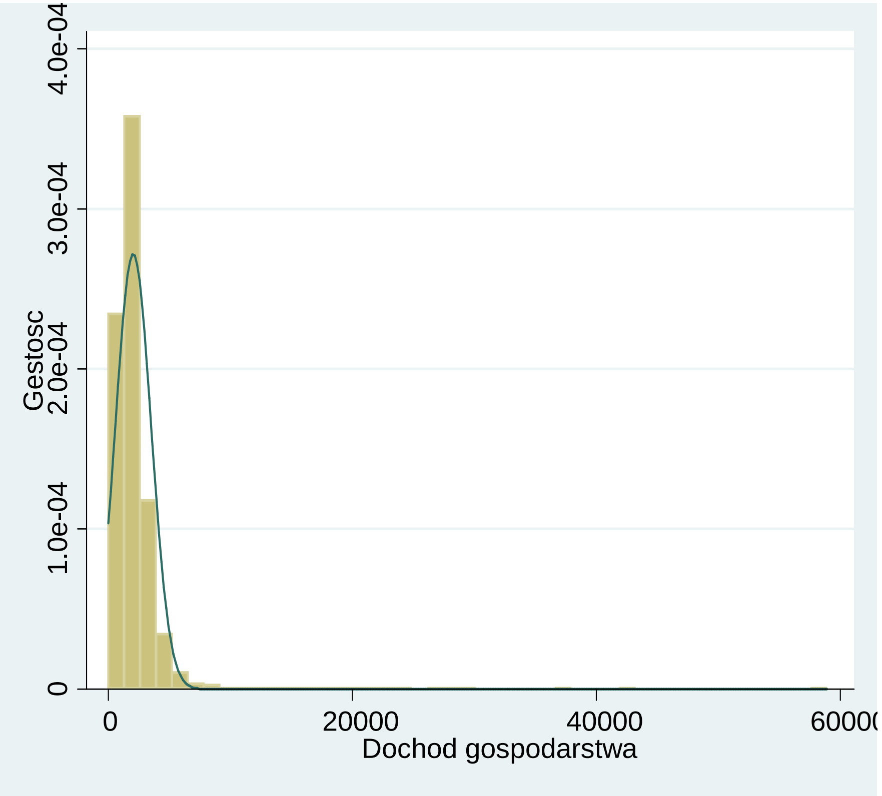
Modelujemy wydatki gospodarstw domowych za pomocą dochodu tych gospodarstw.

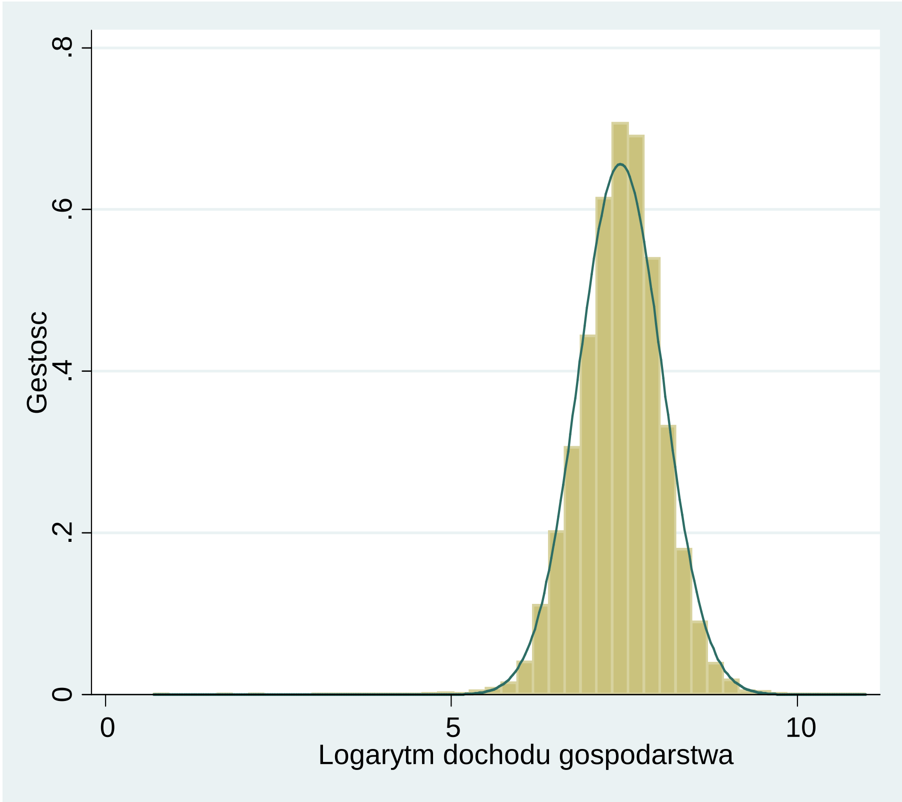
Histogram wydatków /logarytmu wydatków gospodarstw domowych:



Zastosowanie modelu potęgowego - przykład

Histogram dochodów/logarytmu dochodów gospodarstw:





Zastosowanie modelu potęgowego - przykład

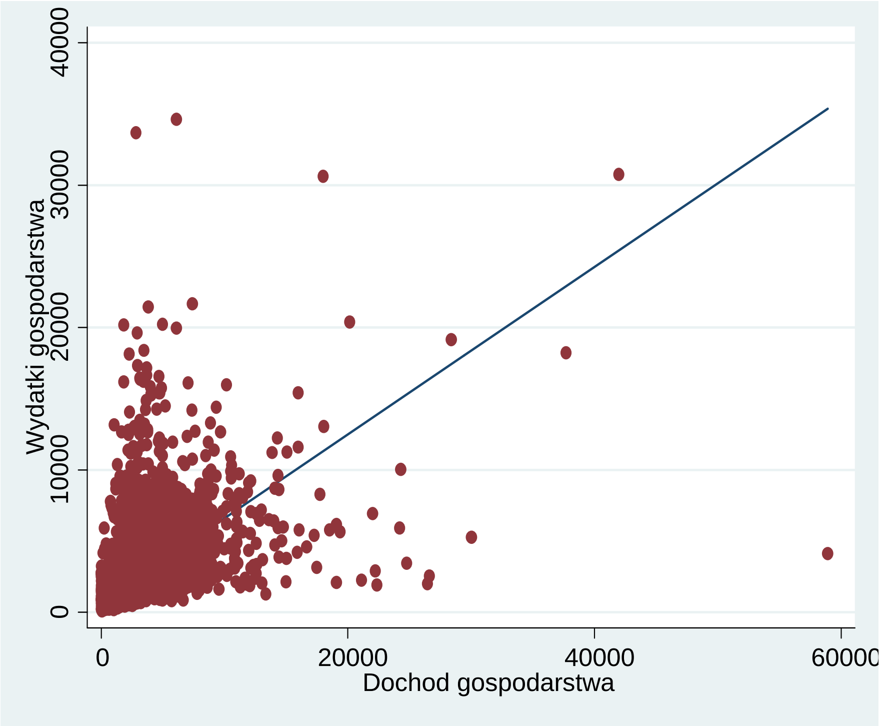
Wyniki regresji:

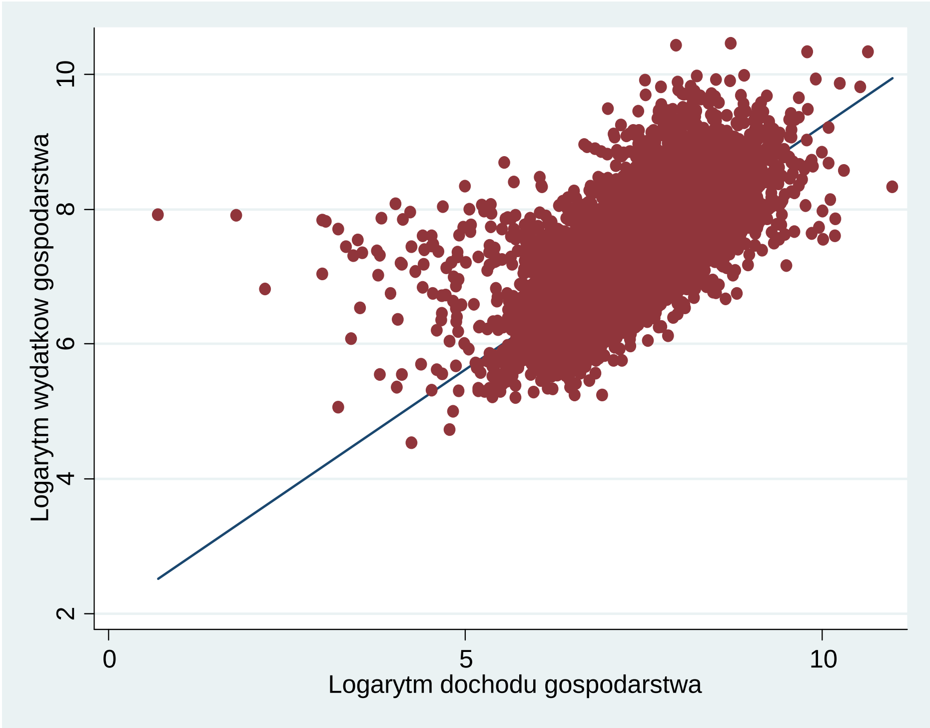
$\ln(Wydatki) = 2{,}02+0{,}72*\ln(Dochod)$ $R^2=0{,}58$

$Wydatki=712{,}81+0{,}58*Dochod$ $R^2=0{,}41$

Zastosowanie modelu potęgowego - przykład

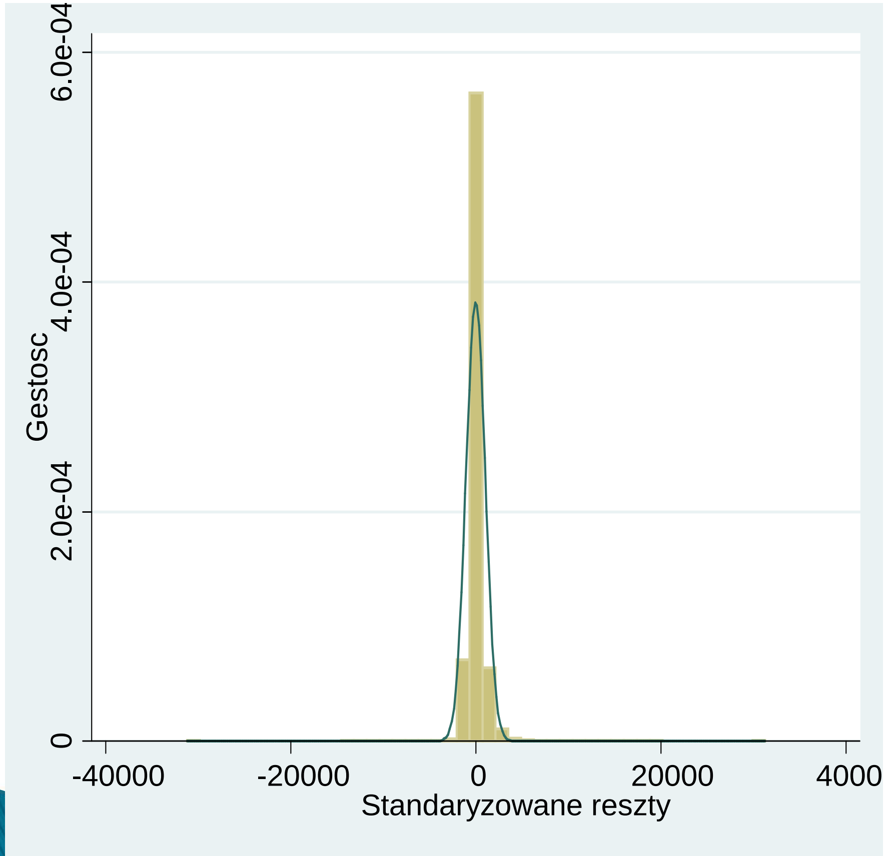
Regresja na poziomach i logarytmach:

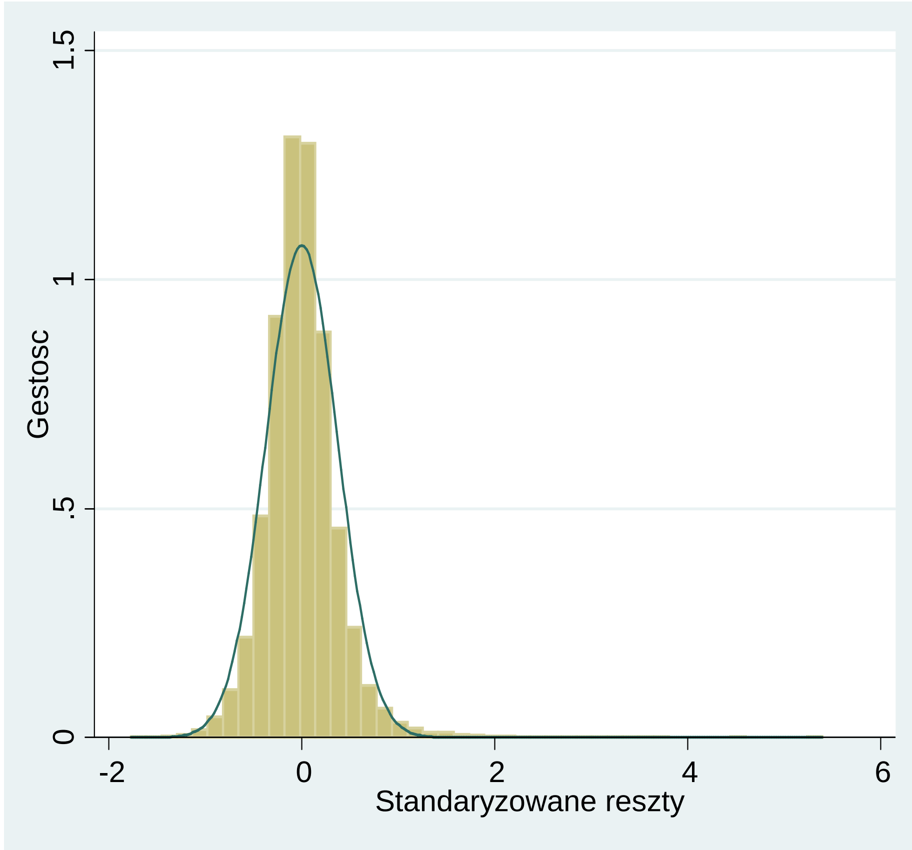




Zastosowanie modelu potęgowego - przykład

Reszty z regresji:





Plan wykładu

- 1. Zastosowanie modelu potęgowego
  - Przekształcenie Boxa-Coxa
- 2. Zmienne ciągłe za zmienne dyskretne
- 3. Interpretacja parametrów przy zmiennych dyskretnych

Przekształcenie Boxa-Coxa

- Pozwala na przeprowadzenie sformalizowanej procedury wyboru między modelem liniowym i potęgowym

- Postać przekształcenia:

$$x^{(\lambda)} = g(x, \lambda) = \frac{x^{\lambda} - 1}{\lambda}$$

Pytanie: Co otrzymujemy dla $\lambda = 1$, $\lambda = -1$, $\lambda = 0$?

Przekształcenie Boxa-Coxa

- Stosując to przekształcenie do zmiennej zależnej i zmiennych niezależnych:

$$y_i^{(\lambda)} = \beta_1 + x_{2i}^{(\lambda)}\beta_2 + \ldots + x_{Ki}^{(\lambda)}\beta_K + \varepsilon_i$$

- Dla $\lambda = 1$:

$$y_i = \beta^*{}_1 + x_{2i}\beta_2 + \ldots + x_{Ki}\beta_K + \varepsilon_i$$

Gdzie: $\beta_1^* = 1 + \beta_1 - \beta_2 - \ldots\beta_K$

Przekształcenie Boxa-Coxa

- Dla $\lambda = -1$:

$$\frac{1}{y_i} = \beta^*{}_1 + \frac{\beta_2}{x_{2i}} + \ldots + \frac{\beta_K}{x_{Ki}} + \varepsilon_i$$

Gdzie: $\beta_1^* = 1 - \beta_1 - \beta_2 - \ldots \beta_K$

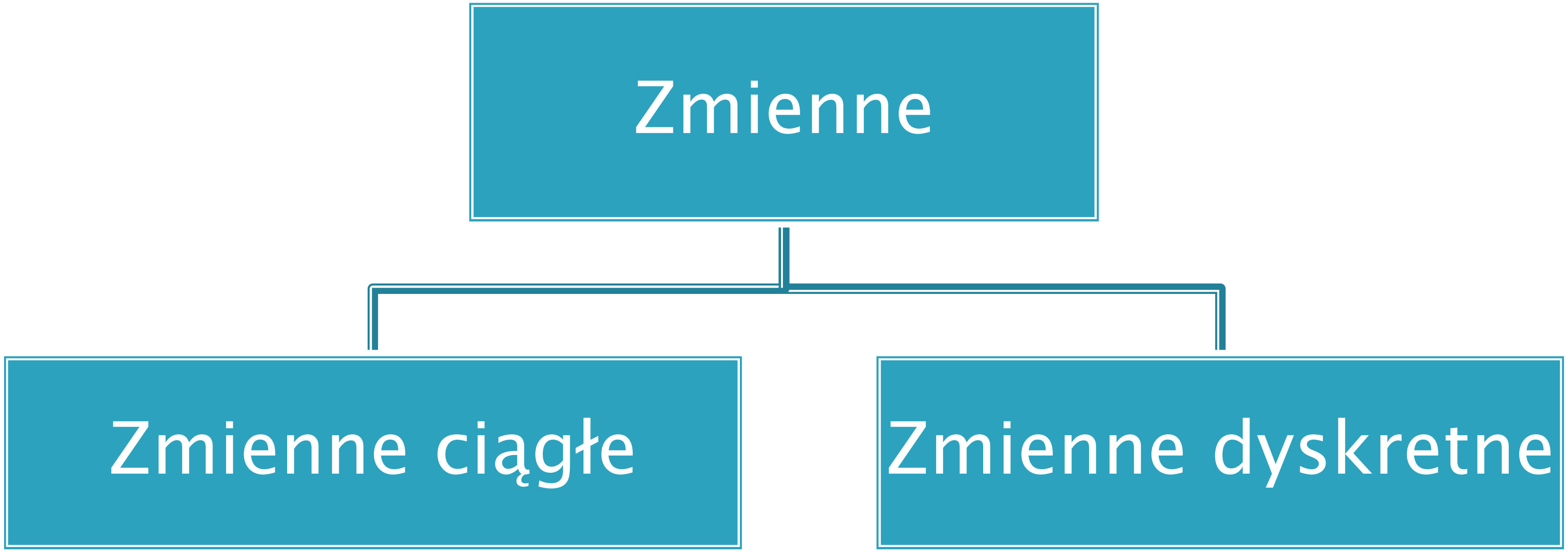
Dla $\lambda = 0$:

$$\ln y_i = \beta_1 + \ln x_{2i}\beta_2 + \ldots + \ln x_{Ki}\beta_K + \varepsilon_i$$

Plan wykładu

- 1. Zastosowanie modelu potęgowego
  - Przekształcenie Boxa-Coxa
- 2. Zmienne ciągłe za zmienne dyskretne
- 3. Interpretacja parametrów przy zmiennych dyskretnych

Zmienne



Zmienne ciągłe

- Zmienną ciągłą nazywamy zmienną, która przyjmuje wartości ze zbioru liczb rzeczywistych.
- Zmienne ciągłe są zmiennymi posiadającymi charakter ilościowy
- Np. dochody, wydatki, cena nieruchomości itd.

Zmienne dyskretne

- Zmienną dyskretną nazywamy zmienną, która przyjmuje wartości ze skończonego podzbioru liczb naturalnych.
- Zazwyczaj podzbiór ten jest stosunkowo mało liczny – obejmuje kilka czy kilkanaście elementów.
- Zmienne dyskretne są zmiennymi posiadającymi charakter jakościowy.
- np. płeć, wykształcenie, miejsce zamieszkania, stan cywilny i itd.

Plan wykładu

- 1. Zastosowanie modelu potęgowego
  - Przekształcenie Boxa-Coxa
- 2. Zmienne ciągłe za zmienne dyskretne
- 3. Interpretacja parametrów przy zmiennych dyskretnych

Zmienne 0 - 1

- Zmienną zero-jedynkową nazywamy zmienną, która przyjmuje tylko dwie wartości: 0 lub 1

- płeć: 1 – kobieta, 0 – mężczyzna
- praca: 1 – pracujący, 0 – niepracujący
- obecność dzieci: 1 – nie, 0 – tak

- Uwaga!
- Ważne jest, że zmienna przyjmuje dwie wartości, nie ma znaczenia ich wielkość.

Interpretacja przy zmiennej 0 – 1 w modelu liniowym względem zmiennych objaśniających

$$\mathrm{D}_i = \begin{cases} 1 & \text{poziom badany} \\ 0 & \text{poziom bazowy} \end{cases}$$

- Niech D_i będzie zmienną zero-jedynkową:

$$y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \gamma D_i + \varepsilon_i$$

- Dla D_i =1 model ma postać:

$$y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \gamma + \varepsilon_i$$

- Dla D_j =0 model ma postać:

$$y_j = \beta_1 + \beta_2 X_{2j} + \ldots + \beta_K X_{Kj} + \varepsilon_j$$

- Zatem $\gamma = \mathrm{E}(y_i) - \mathrm{E}(y_j)$

Interpretacja przy zmiennej 0 – 1 w modelu liniowym względem zmiennych objaśniających

- Wniosek:

- Wielkość γ można interpretować jako zmianę oczekiwanej wartości ***y***, jeśli D zmieni się z 0 na 1, przy założeniu pozostałych charakterystyk na niezmienionym poziomie.

Przykład

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \gamma D_i + \varepsilon_i$$

$$\hat{Y}_i = b_1 + b_2 X_{2i} + \ldots + b_K X_{Ki} + \hat{\gamma} D_i$$

γ – współczynnik przy zmiennej 0-1

INTERPRETACJA: wartość zmiennej zależnej y dla poziomu zmiennej 0-1 D=1 jest:

- większa (jeżeli $\hat{\gamma}$ >0) o | $\hat{\gamma}$ | jednostek lub
- mniejsza (jeżeli $\hat{\gamma}$ <0) o | $\hat{\gamma}$ | jednostek

niż wartość zmiennej zależnej y dla poziomu zmiennej 0-1 D=0 (dla poziomu bazowego)

Przykład

$$placa_i = \beta_1 + \beta_2 plec_i + \varepsilon_i$$

$$\widehat{placa}_i = 926{,}1 - 503{,}59 \cdot plec_i$$

- Zmienna $\quad \text{plec}_i = \begin{cases} 1 & \text{jesli kobieta} \\ 0 & \text{jesli mezczyzna} \end{cases}$
- Interpretacja:

  Oczekiwany poziom płac kobiet jest średnio o 503, 59 złotego niższy niż dla mężczyzn, przy założeniu pozostałych charakterystyk na niezmienionym poziomie.

Przykład

$$placa_i = \beta_1 + \beta_2 sex_i + \varepsilon_i$$

$$\widehat{placa}_i = 422,51 + 503,59 \cdot sex_i$$

- Zmienna $sex_i = \begin{cases} 1 \text{ jesli mezczyzna} \\ 0 \quad \text{jesli kobieta} \end{cases}$
- Interpretacja:

  Oczekiwany poziom płac mężczyzn jest średnio o 503, 59 złotego wyższy niż dla kobiet, przy założeniu pozostałych charakterystyk na niezmienionym poziomie.

Interpretacja przy zmiennej 0 – 1
w modelu, którym zmienna zależna jest zlogarytmowana

$$\ln Y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki} + \gamma D_i + \varepsilon_i$$

- Wniosek:

Wielkość γ (przemnożoną przez 100%)
można interpretować jako procentową zmianę oczekiwanej wartości zmiennej zależnej ***y***, jeśli D zmieni się z 0 na 1 .

Przykład

$$\ln(placa_i) = \beta_1 + \beta_2 plec_i + \varepsilon_i$$

$$\widehat{\ln(placa_i)} = 7{,}67 - 0{,}17 \cdot plec_i$$

- Zmienna $\mathrm{plec}_i = \begin{cases} 1 & \text{jesli kobieta} \\ 0 & \text{jesli mezczyzna} \end{cases}$
- Interpretacja:

  Oczekiwany poziom płac kobiet jest średnio o 17% niższy niż dla mężczyzn, przy założeniu pozostałych charakterystyk na niezmienionym poziomie.

Zmienne dyskretne

- Nieco bardziej skomplikowana jest sytuacja, gdy mamy do czynienia ze zmienną dyskretną która przyjmuje więcej niż 2 wartości.
- np. **wykształcenie** (1 – podstawowe, 2 – średnie, 3 - wyższe)
- W tym przypadku do każdego poziomu **s** zmiennej dyskretnej $\mathbf{X_i}$ musimy przypisać jedną zmienną zero-jedynkową **Ds,i**

$$D_{s,i} = 1 \text{ gdy } X_i = s$$
$$D_{s,i} = 0 \text{ gdy } X_i \neq s \text{ dla } s = 1,2,\ldots,S$$

Przykład

$$\text{wykształcenie}_i = \begin{cases} 1 & \text{podstawowe} \\ 2 & \text{średnie} \\ 3 & \text{wyzsze} \end{cases}$$

$$\text{podstawowe}_i = \begin{cases} 1 & \text{podstawowe} \\ 0 & \text{w p. p.} \end{cases}$$

$$\text{średnie}_i = \begin{cases} 1 & \text{średnie} \\ 0 & \text{w p. p.} \end{cases}$$

$$\text{wyzsze}_i = \begin{cases} 1 & \text{wyzsze} \\ 0 & \text{w p. p.} \end{cases}$$

Zmienne dyskretne

- Za poziom bazowy uznajemy jeden z poziomów (np. poziom 1), i zmienną zero-jedynkową związaną z tym poziomem usuwamy z modelu ze stałą.

- Np. dla zmiennej wykształcenie
- Poziom bazowy : wykształcenie podstawowe

$$placa_i = \beta_1 + \beta_2 średnie_i + \beta_3 wyzsze_i + \varepsilon_i$$

- Dlaczego?

  Nie jest możliwe, by w modelu była jednocześnie stała i wszystkie zmienne zero-jedynkowe (dla każdego poziomu zmiennej dyskretnej), ponieważ macierz $X^T X$ byłaby osobliwa!

Interpretacja parametrów przy zmiennych dyskretnych

- Interpretacja współczynników w modelu z wieloma zmiennymi 0-1 (zmiennymi dyskretnymi) jest analogiczna jak w przypadku modelu z jedną tylko taką zmienną:

  dany współczynnik opisuje różnicę między oczekiwaną wartością zmiennej **y** dla respondenta o charakterystyce bazowej i dla respondenta o charakterystyce s.

Przykład

Modelujemy płace za pomocą płci, wieku i wykształcenia:

Zmienna	Współczynniki
Płeć	−0,278
Wiek	0,078
Wykszt. średnie	−0,273
Wykszt. średnie zawodowe	−0,273
Wykszt. zawodowe	−0,444
Wykszt. podstawowe	−0,571
Stała	6,64

Dlaczego rozkodowujemy zmienne dyskretne

$$\ln(placa_i) = \beta_1 + \beta_2 plec_i + \beta_3 wyksztalcenie + \varepsilon_i$$

$$\ln(placa_i) = \beta_1 + \beta_2 plec_i + \beta_3 wyksztalcenie + \beta_4 wojewodztwo + \varepsilon_i$$

Dziękuję za uwagę