

ZMIENNE BINARNE, KATEGORYCZNE, TRANSFORMACJA LOGARYTMICZNA.

ZAJĘCIA NR 8.

Wszystkie ćwiczenia pochodzą z podręczników Wooldridge lub Stock & Watson.

1. (Na podstawie Wooldridge, 2 s.256) Poniższe modele oszacowano używając danych BWGHT:

$$\log(bwght) = 4.66 - .0044cigs + .0093\log(faminc) + .016parity + .027male + .055white$$

$n = 1388$, $R^2 = 0.0472$, oraz:

$$\begin{aligned} \log(bwght) = 4.65 - .0052cigs + .0110\log(faminc) + .017parity + .034male + .045white \\ - .0030motheduc + .0032fatheduc \end{aligned}$$

$n = 1191$, $R^2 = 0.0493$.

Opis zmiennych:

- $bwght$ = waga urodzenia (funty)
- $cigs$ = średnia liczba papierosów wypalonych dziennie podczas ciąży
- $parity$ = kolejność urodzenia dziecka
- $faminc$ = roczny dochód rodziny
- $motheduc$ = lata edukacji matki
- $fatheduc$ = lata edukacji ojca
- $male$ = dziecko jest płci męskiej
- $white$ = dziecko jest rasy białej

(i) W pierwszym modelu, zinterpretuj współczynnik przy zmiennej $cigs$. W szczególności, jaki jest efekt na wagę przy urodzeniu 10-ciu więcej wypalonych papierosach dziennie?

(ii) O ile większa jest przewidywana waga białego dziecka w porównaniu z pozostałymi dziećmi, przy innych czynnikach niezmiennych w pierwszym modelu?

(iii) Omów szacowany efekt zmiennej $motheduc$.

2. (Na podstawie Wooldridge, 4 s.259) Mamy dany model objaśniający wynagrodzenie CEO:

$$\log(salary) = 4.59 + .257\log(sales) + .011roe + .158finance + .181consprod - .283utility + u$$

$n = 5209$, $R^2 = 0.357$. Użyto danych CEOSAL1, gdzie $finance$, $consprod$, oraz $utility$ to zmienne binarne wskazujące sektory finansowy, produktów konsumenckich, oraz użyteczności publicznej. Pominięta kategoria to sektor transportowy.

(i) Oblicz przybliżoną procentową różnicę w oszacowanej płacy pomiędzy sektorami użyteczności publicznej oraz transportowym, przy innych czynnikach niezmiennych.

(ii) Oblicz dokładną zmianę procentową w szacowanej płacy między sektorami użyteczności publicznej i transportowym, oraz porównaj tę odpowiedź z podpunktem (i).

(iii) Jaka jest przybliżona procentowa zmiana w szacowanej płacy pomiędzy sektorami produktów konsumenckich oraz finansowym?

3. Użyj danych **NBASAL** w tym ćwiczeniu (Wooldridge, C10 s.265).

- (i) Oszacuj model objaśniający punkty na mecz (*points per game*) doświadczeniem oraz pozycją (obrońca, skrzydłowy oraz center). Uwzględnij doświadczenie podniesione do kwadratu oraz użyj centrów jako grupy odniesienia. Jakie są wyniki?
- (ii) Dlaczego nie uwzględnisz wszystkich zmiennych binarnych opisujących pozycje w podpunkcie (i)?
- (iii) Przy niezmiennym doświadczeniu, czy wynik obrońcy jest większy od wyniku centra? O ile więcej?
- (iv) Teraz dodaj zmienną opisującą status matrymonialny do modelu. Czy przy niezmienionej pozycji oraz doświadczeniu, żonaci gracze są bardziej efektywni (na podstawie punktów na mecz)?
- (v) Oszacuj model z podpunktu (iv) ale użyj asysty na mecz jako zmienną zależną. Czy zauważasz różnice w porównaniu z wynikami z podpunktu (iv)? Omów.

4. Użyj danych **GPA2** w tym ćwiczeniu (na podstawie Wooldridge, C4 s.262).

(i) Mamy dany model:

$$colgpa = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hspc + \beta_4 sat + \beta_5 female + \beta_6 athlete + u$$

gdzie *colgpa* to skumulowana średnia ocen, *hsize* to rozmiar klasy absolwentów (w setkach), *hspc* percentyl w klasie absolwentów, *sat* to wynik testu SAT, *female* to zmienna zero-jedynkowa oraz *athlete* to zmienna zerojdynkowa oznaczająca studenta-sportowca. Jakich oszacowań się spodziewasz? Co do których nie jesteś pewien/pewna?

- (ii) Oszacuj model z podpunktu (i). Jaka jest różnica w średniej ocen między sportowcami i pozostałymi?
- (iii) Usuń *sat* z modelu i oszacuj ponownie równanie. Jaka jest szacowany efekt bycia sportowcem? Omów dlaczego oszacowanie jest różne od tego z podpunktu (ii).
- (iv) Zmień model z podpunktu (i), tak aby sprawdzić czy efekt bycia sportowcem różni się dla mężczyzn i kobiet.

5. Użyj danych **firms2000** w tym ćwiczeniu. Oszacujemy funkcję produkcji, jako przykład zastosowania modelu potęgowego. Przypomnijmy:

$$y = k^\alpha l^\beta$$

po transformacji logarytmicznej:

$$\log(y) = \alpha \log(k) + \beta \log(l)$$

równanie regresji:

$$\log(y) = \beta_0 + \alpha k + \beta l + u$$

gdzie *y* to produkcja *k* to nakład kapitału a *l* to nakład pracy.

(i) Używając dostępnych miar ze zbioru danych, oszacuj poniższe równanie

$$AV = \beta_0 + \alpha TOAS + \beta EMPL + u$$

gdzie AV to wartość dodana (tys. PLN), $TOAS$ - całkowity kapitał firmy (tys. PLN), $EMPL$ - wielkość zatrudnienia (liczba pracowników). Zinterpretuj współczynniki.

- (ii) Do równania z (i) dodaj zmienną zero-jedynkową przyjmującą wartość 1 dla przemysłu i 0 dla usług (należy ją utworzyć korzystając ze zmiennej *industry_factor* oraz funkcji `ifelse()`). Oszacuj nowy model oraz zinterpretuj nowy współczynnik.
- (iii) Dodaj do modelu z (i) zmienną przyjmującą wartość 1 jeśli firma zatrudnia więcej niż 100 pracowników. Zinterpretuj współczynniki. Czy między tymi grupami są duże różnice?
- (iv) (*) Oszacuj model z (i) lecz pozwól różnić się efektom rozmiaru między sektorami. Omów wyniki. (Stwórz zmienne binarne opisujące duże firmy w przemyśle, małe firmy w przemyśle, i podobnie w usługach... Pamiętaj o wybraniu poziomu odniesienia!)