

# ZMIENNE BINARNE, INTERAKCJE I MODELE Z NIELINIOWOŚCIĄ.

## I. Zmienne binarne & kategoriowe.

1. Mamy dane modele objaśniające wynagrodzenie CEO:

$$salary = 943.96 + 0.013sales + 4.99roe + 560.32consprod - 320.85utility + 253.46finance + e \quad (1)$$

$$\log(salary) = 4.59 + 0.257\log(sales) + 0.011roe + 0.181consprod - 0.283utility + 0.158finance + e \quad (2)$$

$n = 5209$ ,  $R^2 = 0.357$ . gdzie *finance*, *consprod*, *utility*, *indus* to zmienne binarne wskazujące sektory finansowy, produktów konsumenckich, użyteczności publicznej oraz sektor transportowy.

- (i) Oblicz przybliżoną (procentową) różnicę w oszacowanej płacy pomiędzy sektorami użyteczności publicznej oraz transportowym, przy innych czynnikach niezmiennych dla obu oszacowań.
  - (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?
2. Oszacowano model objaśniający zdobywane punkty na mecz przez zawodników NBA (*points*) doświadczeniem oraz zmiennymi binarnymi dla każdej z trzech pozycji (obrońca (*guard*), skrzydłowy (*forward*) oraz center(*center*)).

$$points = 4.76 + 2.31guard + 1.54forward + 1.28exper - 0.071exper^2 + e \quad (3)$$

- (i) Dlaczego nie uwzględniono wszystkich zmiennych binarnych opisujących pozycję?
- (ii) Przy niezmiennym doświadczeniu, czy wynik obrońcy jest większy od wyniku centra? O ile więcej?
- (iii) Dodano do modelu zmienną opisującą status matrymonialny (*marr*) zawodnika.

$$points = 4.70 + 2.28guard + 1.54forward + 1.23exper - 0.07037exper^2 + 0.58427marr + e$$

Czy przy niezmiennym doświadczeniu, żonaci gracze są bardziej efektywni (na podstawie punktów na mecz)?

- (iv) Oszacowano model z poprzedniego podpunktu ale teraz zmienną objaśnianą są asysty na mecz (*assists*).

$$assists = -0.23 + 2.49guard + 0.45forward + 0.44exper - 0.028exper^2 + 0.32marr + e \quad (4)$$

Czy zauważasz różnice w porównaniu z wynikami z poprzednich podpunktów? Omów.

## II. Modele z wielomianami.

1. Oszacowano równanie objaśniające płace:

$$\log(wage) = 0.128 + 0.09educ + 0.041exper - 0.00071exper^2 + e$$

- (i) Używając przybliżenia, znajdź zwrot z piątego roku doświadczenia. Jaki jest przybliżony zwrot z dwudziestego roku doświadczenia?
- (ii) Przy jakiej wartości *exper*, zwiększenie doświadczenia obniża przewidywany logarytm płacy? (Hint: Dla jakiej wartości doświadczenia, płaca jest największa?)

2. Oszacowano równanie objaśniające logarytm wagi noworodka:

$$\log(bwght) = 7.957 + 0.0189npvis - 0.00043npvis^2 + e$$

- (i) Pokaż korzystając z oszacowania, że liczba wizyt prenatalnych (*npvis*) która maksymalizuje  $\log(bwght)$  wynosi około 22.
- (ii) Czy ma to sens, że waga urodzenia spada po przekroczeniu 22 wizyt prenatalnych? Wyjaśnij.
- (iii) Dodano do modelu zmienną opisującą wiek matki do równania oraz jej kwadrat:

$$\log(bwght) = 7.58 + 0.0253mage - 0.00041mage^2 + 0.018npvis - 0.00041npvis^2 + e$$

Przy *npvis* ustalonym, czy jakim wieku matki waga przy narodzinach dziecka jest największa?

## III. Modele z interakcjami

1. Oszacowano równanie objaśniające wynik testu:

$$sat = 1028.097 + 19.297hsize - 2.195hsize^2 - 45.091female - 169.813black + 62.306female \times black + e$$

Zmienna *sat* to łączny wynik testu SAT, *hsize* to wielkość klasy absolwentów (w setkach), *female* to zmienna zero-jedynkowa oraz *black* to zmienna zero-jedynkowa oznaczająca rasę.

- (i) Korzystając z oszacowania wyznacz optymalną wielkość klasy.
- (ii) Przy niezmiennym *hsize*, jaka jest szacowana różnica w wyniku SAT między nieczarnymi kobietami i czarnymi mężczyznami?
- (iii) Jaka jest szacowana zmiana w wyniku SAT między czarnymi mężczyznami a pozostałymi mężczyznami?
- (iv) Jaka jest szacowana różnica między czarnymi kobietami a pozostałymi kobietami?

2. Oszacowano równanie objaśniające płace:

$$\log(wage) = 2.28877 - 0.35727female + 0.04966totcoll + 0.02991female \times totcoll$$

- (i) Używając oszacowań, znajdź wartości *totcoll* (lata w college) takie że przewidywane wartości  $\log(wage)$  są takie same dla kobiet i mężczyzn.
- (ii) Korzystając z oszacowania powiedz, czy kobiety naprawdę mogą uzyskać tyle lat edukacji, aby ich zarobki dogoniły zarobki mężczyzn? Wyjaśnij.

3. Oszacowano poniższy model, który wiąże zwrot z edukacji z edukacją rodziców (*pareduc*):

$$\log(wage) = 5.646519 + 0.046752educ + 0.018871exper + 0.010217tenure + 0.000775educ \times pareduc + e$$

- (i) w przybliżeniu zwrot z dodatkowego roku edukacji można opisać formułą:

$$\frac{\Delta \log(wage)}{\Delta educ} = \beta_1 + \beta_4 \text{pareduc}$$

Jakiego znaku spodziewasz się przy  $\beta_4$

- (ii) Zinterpretuj współczynnik przy interakcji. Pomocnym może być wybranie dwóch wartości zmiennej *pareduc* np.: *pareduc* = 32 (obaj rodzice mają wykształcenie wyższe), lub *pareduc* = 24 (oboje rodziców mają wykształcenie średnie) i porównaj oszacowany zwrot z edukacji.
- (iii) Dodano *pareduc* jako oddzielną zmienną i oszacowano model:

$$\log(wage) = 4.937 + 0.0971educ + 0.0196exper + 0.0103tenure + 0.033pareduc - 0.0016educ \times pareduc + e$$

Czy teraz zależność między zwrotem z edukacji a edukacją rodziców jest pozytywna? Zinterpretuj współczynnik przy interakcji.

4. Mamy dany model:

$$\text{colgpa} = \beta_0 + \beta_1 \text{hsize} + \beta_2 \text{hsize}^2 + \beta_3 \text{hsperc} + \beta_4 \text{sat} + \beta_5 \text{female} + \beta_6 \text{athlete} + \varepsilon$$

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

$$\text{colgpa} = 1.241 - 0.0569\text{hsize} + 0.00466\text{hsize}^2 - 0.013213\text{hsperc} + 0.00165\text{sat} + 0.1549\text{female} + 0.169\text{athlete} + e$$

Jaka jest różnica w średniej ocen między sportowcami i pozostałymi?

- (iii) Zapisz równanie modelu taki sposób, aby sprawdzić czy efekt bycia sportowcem różni się dla mężczyzn i kobiet.