

REGRESJA LINIOWA - ZADANIA.

ZAJĘCIA NR 6.

1. Mamy trzy modele

(i) $y = \beta_0 + \beta_1 x_1 + \varepsilon$

(ii) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

(iii) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

Jaka zachodzi relacja między RSS w regresjach? Jaka zachodzi relacja między R^2 w tych regresjach?

2. Na podstawie danych WAGE2 . RAW o pracujących mężczyznach uzyskano poniższe oszacowanie:

$$\widehat{educ} = 10.36 - 0.094 \times sibs + 0.131 \times meduc + 0.210 \times feduc$$

$$n = 5.722, R^2 = 0.214$$

gdzie *educ* to lata edukacji, *sibs* oznacza liczbę rodzeństwa, *meduc* to lata edukacji matki, *feduc* to lata edukacji ojca.

(i) Czy rodzeństwo ma efekt zgodny z oczekiwaniami? Wyjaśnij. Przy niezmiennych *meduc* oraz *feduc*, o ile musi zmienić się *sibs* aby przewidywane lata edukacji obniżyły się o rok?

(ii) Omów interpretację współczynnika przy zmiennych *meduc*.

(iii) Przyjmijmy że mężczyzna A nie ma rodzeństwa, a jego rodzice mają po 12 lat edukacji. Mężczyzna B nie ma rodzeństwa, a jego rodzice mają po 16 lat edukacji każdy. Jaka jest przewidywana zmiana różnica w latach edukacji między mężczyznami A i B?

3. Poniższy model to uproszczona wersja modelu regresji wielorakiej używanego przez Biddle i Hamermesh'a (1990) do studiowania relacji między czasem poświęconym na sen oraz pracę, przy innych czynnikach wpływających na sen:

$$sleep = \beta_0 + \beta_1 totwork + \beta_2 educ + \beta_3 age + u$$

gdzie *sleep* oraz *totwork* są mierzone w minutach na tydzień, a *educ* oraz *age* są mierzone w latach.

(i) Jeśli dorośli śpią mniej aby pracować, jaki będzie znak β_1 ?

(ii) Jak myślisz, jakie znaki będą miały β_2 oraz β_3 ?

(iii) Używając danych SLEEP75 . RAW otrzymano poniższe oszacowanie:

$$\widehat{sleep} = 3,638.25 - 0.148 totwork - 11.13 educ + 2.20 age$$

$$n = 5,706, R^2 = 0.113$$

Jeżeli ktoś pracuje 5 godzin więcej tygodniowo, o ile minut spadnie przewidywany czas snu? Czy to dużo?

(iv) Omów znak oraz wielkość oszacowanego współczynnika przy zmiennej *educ*.

(v) Czy twoim zdaniem zmienne *totwork*, *educ* oraz *age* wyjaśniają dużo zmienności zmiennej *sleep*? Czy inne czynniki mogą być związane z długością czasu poświęcanego na sen? Czy te czynniki są związane z *totwork*?

4. Dane zebrano z losowej próby 220 sprzedanych domów z pewnej miejscowości w 2013 roku. Niech *Price* oznacza cenę sprzedaży (w \$1000), *BDR* oznacza liczbę sypialni, *Bath* oznacza liczbę łazienek, *Hsize* oznacza

rozmiar domu (w mkw.), $Lsize$ oznacza rozmiar działki (w mkw.), Age oznacza wiek domu w latach, $Poor$ oznacza zmienną zerojedynkową, która przyjmuje wartość 1 jeżeli stan domu jest słaby. Oszacowano regresję objaśniającą cenę domu:

$$Price = 119.2 + 0.485BDR + 23.4Bath + 0.156Hsize + 0.002Lsize + 0.090Age - 48.8Poor$$

$$\bar{R}^2 = 0.72, SER = 41.5.$$

- (i) Przyjmijmy że właściciel przerobi część pokoju dziennego na jedną łazienkę. Jaki jest oczekiwany wzrost wartości domu?
 - (ii) Przypuśćmy że właściciel doda nową łazienkę do swojego domu, co zwiększy jego rozmiar o 100 mkw. Jaki jest oczekiwany wzrost wartości domu?
 - (iii) Jaki będzie spadek wartości domu, jeżeli właściciel zaniedba go, tak że jego stan będzie słaby?
 - (iv) Oblicz R^2 i zinterpretuj go.
5. Używając danych KIELMC tylko dla roku 1981, odpowiedz na poniższe pytania. Dane opisują domy sprzedane w 1981 w North Andover, Massachusetts. W 1981 roku rozpoczęła się budowa lokalnej spalarni śmieci.
- (i) Aby zbadać wpływ lokalizacji spalarni na ceny domów, zbudowano prosty model:

$$\log(price) = \beta_0 + \beta_1 \log(dist) + u$$

gdzie $price$ to cena domu w dolarach a $dist$ to dystans domu do spalarni, mierzony w stopach. Jakiemu znaku spodziewasz się przy β_1 jeżeli obecność spalarni obniża ceny domów? Oszacuj model oraz zinterpretuj wyniki.

- (ii) Do prostego modelu z (i), dodano zmienne $\log(intst)$, $\log(area)$, $\log(land)$, $rooms$, $baths$ oraz age , gdzie $intst$ to dystans domu od autostrady, $area$ to powierzchnia domu w mkw., $land$ to rozmiar działki w mkw., $rooms$ to całkowita liczba pokoi, $baths$ to liczba łazienek oraz age to wiek domu w latach. Co teraz możesz powiedzieć o wpływie spalarni? Wyjaśnij dlaczego (i) oraz (ii) dają inne wyniki.
6. Skorzystaj z danych o cenach domów HPRICE1.

- (i) Oszacuj model

$$\log(price) = \beta_0 + \beta_1 \log(lotsize) + \beta_2 \log(sqrft) + \beta_3 bdrms + u$$

oraz zapisz wyniki w formie równania.

- (ii) Znajdź wartości dopasowane $\log(price)$, gdy $lotsize = 20.000$, $sqrft = 2.500$, oraz $bdrms = 4$. Znajdź dopasowane wartości $price$ przy tych samych wartościach zmiennych objaśniających.
- (iii) Aby wyjaśnić zróżnicowanie zmiennej $price$, zdecyduj czy preferujesz model z (i) czy $price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$.