

REGRESJA LINIOWA - ZADANIA.

ZAJĘCIA NR 5.

Wszystkie ćwiczenia pochodzą z podręczników Wooldridge oraz Stock & Watson.

1. Używając danych GPA2 .RAW o 4,137 studentach, otrzymano poniższe oszacowanie MNK:

$$\widehat{colgpa} = 1.392 - 0.0135 \times hsperc + 0.00148 \times sat$$

$$n = 4,137, R^2 = 0.273$$

gdzie $colgpa$ jest mierzony na 4-stopniowej skali, $hsperc$ to percentyl pośród klasy absolwentów szkoły średniej (np. $hsperc = 5$ oznacza top 5% klasy), a sat to łączny wynik testu SAT.

- (i) Dlaczego ujemny współczynnik przy zmiennej $hsperc$ ma sens?
 - (ii) Jaki jest teoretyczny $colgpa$ kiedy $hsperc = 20$ i $sat = 1,050$?
 - (iii) Przyjmij że dwóch absolwentów szkoły średniej A, oraz B, ukończyło szkołę średnią w tym samym percentylu, ale student A uzyskał z testu SAT wynik o 140 punktów wyższy (około jedno odchylenie standardowe). Jaka jest przewidywana różnica w średniej ocen dla tych dwóch studentów? Czy ta różnica jest duża?
 - (iv) Przy innych czynnikach niezmiennych, jaka różnica w wyniku testu SAT, prowadzi do przewidywanej różnicy w średniej ocen na poziomie 0.50 oceny? Skomentuj swoją odpowiedź.
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 zmiennej $meduc$.
 - (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:

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

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

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. Na [stronie](#), znajdziesz plik z danymi `Growth`, który zawiera dane o średnich stopach wzrostu od 1960 do 1995 dla 65 krajów, wraz ze zmiennymi potencjalnie związanymi ze wzrostem. Dokładny opis znajduje się w `Growth_Description`, także dostępny na stronie. W tym ćwiczeniu zbadasz relację pomiędzy wzrostem (`growths`) a handlem (`trade`).

(i) Przygotuj wykres punktowy średniorocznej stopy wzrostu (`Growth`) i średniego udziału handlu (`TradeShare`).

Czy wydaje się że jest związek między zmiennymi?

(ii) Dla Malty, udział w handlu jest o wiele wyższy niż dla innych krajów. Znajdź Maltę na wykresie punktowym. Czy Malta wygląda jak outlier?

(iii) Używając wszystkich obserwacji, wyestymuj regresję `Growth` i `TradeShare`. Jakie jest oszacowane nachylenie? Ile wynosi oszacowanie wyrazu wolnego? Użyj oszacowanej regresji dla kraju z udziałem w handlu na poziomie 0.5 oraz z udziałem w handlu równym 1.0.

(iv) Oszacuj taką samą regresję, ale wyłączając dane dla Malty. Odpowiedź na pytania z (iii).

(v) Narysuj wykres regresji z (iii). oraz (iv). Używając wykresu z (i), wyjaśnij dlaczego prosta regresji jest bardziej stroma niż prosta regresji oszacowana bez Malty?

(vi) Gdzie jest Malta? Dlaczego Malta ma tak duży udział handlu? Czy Malta powinna być wykluczona z analiz?

5. (S&W) Używając danych `GROWTH`, wykluczając dane dla Malty, wykonaj poniższe zadania.

(i) Przygotuj tabelę która prezentuje średnią, odchylenie standardowe, minimalne oraz maksymalne wartości dla zmiennych `Growth`, `TradeShare`, `YearsSchool`, `Oil`, `Rev_Coups`, `Assassinations`, oraz `RGDP60`. Pamiętaj o zawarciu odpowiednich jednostek.

(ii) Oszacuj model objaśniający zmienną `Growth` za pomocą `Trade_Share`, `YearsSchool`, `Rev_Coups`, `Assassinations`, i `RGDP60`. Jaka jest wartość współczynnika przy zmiennej `Rev_Coups`? Zinterpretuj oszacowanie tego współczynnika. Jest duży czy mały, w kategoriach realnego świata?

(iii) Wykorzystaj regresję do przewidywania średniorocznej stopy wzrostu dla kraju ze średnimi wartościami wszystkich zmiennych objaśniających.

(iv) Powtórz (iii) ale załóż, że wartość `TradeShare` dla tego kraju jest jedno odchylenie standardowe ponad średnią.

(v) Dlaczego zmienna `Oil` nie jest uwzględniona w regresji. Co by się stało, gdyby była uwzględniona?

6. (W, C2) Skorzystaj z danych `HPRICE1.RAW` do oszacowania modelu:

$$price = \beta_0 + \beta_1 sqft + \beta_2 bdrms + u$$

gdzie *price* to cena domu w tys, dolarów.

- (i) Zapisz wynik w formie równania.
- (ii) Jaki jest szacowany przyrost ceny domu z dodatkową sypialnią, przy niezmienionej powierzchni?
- (iii) Jaki jest szacowany przyrost ceny domu z dodatkową sypialnią i o powierzchni 140 metrów kw.? Porównaj odpowiedź z odpowiedzią z punktu (ii).
- (iv) Jaki procent zmienności w cenie jest objaśniany powierzchnią (w metrach kw.) oraz liczbą sypialni?
- (v) Pierwszy dom w próbie ma $sqft = 2,438$ oraz $bdrms = 4$. Znajdź przewidywaną cenę dla tego domu, implikowaną przez MNK.
- (vi) Prawdziwa cena pierwszego domu w próbie wynosi \$300,000 (więc $price = 300$). Znajdź resztę dla tego domu. Czy ktoś zapłacił za dużo, czy za mało?

7. Skorzystaj z danych `ATTEND.RAW`.

- (i) Podaj minimum, maksimum oraz średnie wartości dla zmiennych *atndrte*, *priGPA*, oraz *ACT*.
- (ii) Oszacuj model:

$$atndrte = \beta_0 + \beta_1 priGPA + \beta_2 ACT + u$$

i zapisz wynik w formie równania. Zinterpretuj oszacowanie wyrazu wolnego. Czy ma sensowne znaczenie?

- (iii) Omów współczynniki. Czy zauważasz niespodzianki?
- (iv) Jaka jest przewidywana wartość *atndrte* jeżeli $priGPA = 3.65$ oraz $ACT = 20$? Co zrobić z tym wynikiem? Czy są studenci w próbie, z takimi wartościami zmiennych objaśniających?
- (v) Jeśli student A ma $priGPA = 3.1$ i $ACT = 21$ oraz student B ma $priGPA = 2.1$ i $ACT = 26$, to jaka jest przewidywana różnica w ich frekwencji?