NIELINIOWOŚĆ W REGRESJI.

ZAJĘCIA NR 7.

Wszytkie ćwiczenia pochodzą z podręczników Wooldridge oraz Stock & Watson.

1. 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 dollarach a `dist` to dystans domu of spalarni, mierzony w stopach. Jakiego 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 powierchnia 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.
   (iii) Dodaj $(\log(intst))^2$ do modelu z (ii). Co teraz zauważasz. Jak skomentujesz istotność wyboru formy funkcyjnej?
   (iv) Czy dodanie kwadratu $log(dist)$ do modelu z (iii) coś zmienia?

2. Użyj danych `GPA2` do tego zadania.
   (i) Oszacuj model
   $$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize_2 + u$$
   gdzie `hsize` to rozmiar klasy absolwentów (w setkach), oraz zapisz model w postaci równania. Czy zmienna podniesiona do kwadratu jest statystycznie istotna?
   (ii) Używając oszacowań z (i) powiedz jaki jest optymalny rozmiar klasy? Uzasadnij.
   (iii) Czy ta analiza jest reprezentatywnie przedstawia poziom osiągnięć akademickich wśród starszych uczniów liceum? Wyjaśnij.
   (iv) Znajdź optymalny poziom klasy, używając $\log(sat)$ jako zmienną zależną. Czy jest różny od tego otrzymanego w (iii)?

3. 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 dopsowane wartości $price$ przy tych samych wartościach zmiennych objaśniających.
   (iii) Aby wyjaśnić zróżnicowanie zmiennej $price$, zdecyduj czy preferujesz model z (i) czy $price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$.