

3.1 Plik *realest.txt* zawiera następujące dane na temat domów na przedmieściach Chicago: cena domu (*Price*), liczba sypialni (*Bedroom*), powierzchnia w stopach kwadratowych (*Space*), liczba pokoi (*Room*), szerokość frontu działki w stopach (*Lot*), roczny podatek od nieruchomości (*Tax*), liczba łazienek (*Bathroom*), liczba miejsc parkingowych w garażu (*Garage*) i stan domu (*Condition*, 0-dobry, 1-wymaga remontu). Dopasować liniowy model regresji opisujący zależność ceny domu od pozostałych zmiennych w zbiorze.

(b) Wyznacz macierz eksperymentu.

(b) Oblicz estymatory parametrów z definicji i porównaj z wartościami obliczonymi przy użyciu funkcji `coef()`. Oblicz SST, SSR, SSE i współczynnik determinacji z definicji.

(c) Jaki wpływ na cenę ma zwiększenie liczby sypialni o 1, kiedy wartości wszystkich pozostałych zmiennych objaśniających są ustalone? Znaleźć uzasadnienie tego pozornie błędnego wyniku. Porównać ten wynik z wynikiem otrzymanym dla modelu linowego opisującego zależność ceny domu jedynie od liczby sypialni.

(d) Masz dom w tej okolicy, w dobrym stanie, z 3 sypialniami, o powierzchni 1500 stóp kwadratowych, z 8 pokojami, 40 stopami szerokości działki, 2 łazienkami, 1 miejscem w garażu i podatkiem w wysokości 1000 dolarów. Za ile spodziewasz się go sprzedać? Wykonaj predykcje korzystając z definicji oraz funkcji `predict()`.

(e) Oblicz estymator wariancji błędów korzystając z definicji oraz funkcji `summary()`.

3.2 Rozpatrzmy model liniowy.

(a) Udowodnij, że $\sum_{i=1}^n e_i = 0$.

(b) Potwierdź obserwację z pierwszego podpunktu na danych wygenerowanych w następujący sposób:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \varepsilon_i,$$

gdzie $\beta_0 = 2$, $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 0.7$. Niech $x_{i,1}, x_{i,2}, x_{i,3} \sim N(0, 1)$ i $\varepsilon_i \sim N(0, 10)$, a liczba obserwacji $n = 100$.

(c) Powtórz eksperyment z wcześniejszego punktu 1000 razy. Za pomocą otrzymanych wyników wyestymuj wariancję błędu.