

4.1 Zbiór *airpollution.txt* zawiera dane dotyczące związku pomiędzy zanieczyszczeniem powietrza i śmiertelnością w 60 miastach amerykańskich. Między zmiennymi są:

Mortality - skorygowana wiekiem liczba zgonów na 100 000 mieszkańców,

Education - mediana liczby lat kształcenia,

NonWhite - procent tej podpopulacji,

income - mediana zarobków w tys. dolarów,

JanTemp, *JulTemp* - średnie temperatury w styczniu i lipcu (w stopniach Fahrenheita),

NOx - stężenie tlenku azotanu.

(a) Dopasuj model liniowy ze zmienną objaśnianą *Mortality*.

(b) Oblicz z definicji statystykę t i jej p-wartość związaną ze zmienną *NOx*. Sprawdź zgodność z wynikiem dla funkcji `summary()`.

(c) Oblicz z definicji statystykę F i jej p-wartość. Sprawdź zgodność z wynikiem dla funkcji `summary()`.

4.2 Wygeneruj dane w następujący sposób:

$$y_i = \beta_0 + \beta_1 x_{i,1}^2 + \epsilon_i$$

gdzie: $\beta_0 = 0.5$, $\beta_1 = 1$. Niech $x_1 \sim U[0, 1]$ i $\epsilon_i \sim N(0, 1)$.

(a) Dopasuj model **liniowy** (zła specyfikacja!). Oblicz statystykę t i jej p-wartość. Czy zmienna x_1 jest istotna w tym modelu na poziomie istotności 0.05? Ile wynosi wartość statystyki F ?

4.3 Wygeneruj dane w następujący sposób:

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

gdzie: $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.05$. Niech $x_1, x_2, x_3 \sim N(0, 1)$ i $\epsilon_i \sim N(0, 1)$.

(a) Wygeneruj $B = 100$ zbiorów danych z $n = 100$ obserwacji każdy. Na podstawie wygenerowanych zbiorów oszacuj moc testów t dla zmiennych x_1, x_2, x_3 (testy istotności dla poszczególnych zmiennych).

(b) Powtórz powyższy eksperyment dla $n = 20, 50, 100, 200, 300, \dots, 500$. Wykonaj wykres mocy w zależności od liczby obserwacji n (tak aby każda linia odnosiła się do innego współczynnika).

(c) Powtórz eksperyment z punktu (b) dla testu F , dla parametrów $\beta_0 = 0.05$, $\beta_1 = \beta_2 = \beta_3 = 0.05$.