

praca__domowa__3__orchowski

Wojciech Orchowski

Tuesday, May 19, 2015

Regresja liniowa

Model regresji możemy zapisać następująco

$$Y = \beta X + \epsilon$$

Natomiast estymator KMNK β określony jest następująco

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

Dane do regresji

Użyjemy danych ze zbioru cars i zbudujemy model regresji **dist** ~ **speed**. Dodatkowo, losowo (wykorzystując funkcję **sample**) wybierzemy jedynie 20 obserwacji aby dokument był czytelny.

```
car_sub <- cars[sample(1:50,20),]  
speed <- car_sub[,1]  
distance <- car_sub[,2]
```

Macierz X wygląda następująco

$$\mathbf{X} = \begin{bmatrix} 1 & 15 \\ 1 & 10 \\ 1 & 12 \\ 1 & 16 \\ 1 & 12 \\ 1 & 25 \\ 1 & 12 \\ 1 & 15 \\ 1 & 24 \\ 1 & 14 \\ 1 & 20 \\ 1 & 17 \\ 1 & 23 \\ 1 & 4 \\ 1 & 15 \\ 1 & 10 \\ 1 & 13 \\ 1 & 14 \\ 1 & 24 \\ 1 & 17 \end{bmatrix}$$

W związku z tym $\hat{\beta}$ ma postać:

$$(X^T X)^{-1} X^T Y = \left(\begin{bmatrix} 1 & 15 \\ 1 & 10 \\ 1 & 12 \\ 1 & 16 \\ 1 & 12 \\ 1 & 25 \\ 1 & 12 \\ 1 & 15 \\ 1 & 24 \\ 1 & 14 \\ 1 & 20 \\ 1 & 17 \\ 1 & 23 \\ 1 & 4 \\ 1 & 15 \\ 1 & 10 \\ 1 & 13 \\ 1 & 14 \\ 1 & 24 \\ 1 & 17 \end{bmatrix}^T \begin{bmatrix} 1 & 15 \\ 1 & 10 \\ 1 & 12 \\ 1 & 16 \\ 1 & 12 \\ 1 & 25 \\ 1 & 12 \\ 1 & 15 \\ 1 & 24 \\ 1 & 14 \\ 1 & 20 \\ 1 & 17 \\ 1 & 23 \\ 1 & 4 \\ 1 & 15 \\ 1 & 10 \\ 1 & 13 \\ 1 & 14 \\ 1 & 24 \\ 1 & 17 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 15 \\ 1 & 10 \\ 1 & 12 \\ 1 & 16 \\ 1 & 12 \\ 1 & 25 \\ 1 & 12 \\ 1 & 15 \\ 1 & 24 \\ 1 & 14 \\ 1 & 20 \\ 1 & 17 \\ 1 & 23 \\ 1 & 4 \\ 1 & 15 \\ 1 & 10 \\ 1 & 13 \\ 1 & 14 \\ 1 & 24 \\ 1 & 17 \end{bmatrix}^T \begin{bmatrix} 54 \\ 18 \\ 14 \\ 32 \\ 20 \\ 85 \\ 28 \\ 26 \\ 120 \\ 60 \\ 32 \\ 32 \\ 54 \\ 10 \\ 20 \\ 26 \\ 34 \\ 26 \\ 92 \\ 40 \end{bmatrix} = \begin{bmatrix} -25.0323275862068 \\ 4.24245689655172 \end{bmatrix}$$

Co daje nam

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \left(\begin{bmatrix} 20 & 312 \\ 312 & 5424 \end{bmatrix} \right)^{-1}$$

Rysujemy wykres regresji

Wykres przedstawia zależność między **dist** a **speed** oraz oszacowaną krzywą regresji. Należy użyć funkcji **plot** oraz **abline**.

Zależność między prędkością, a dystansem

