

Zadanie dodatkowe - R Markdown

Anna Laskowska

Tuesday, May 19, 2015

Regresja liniowa

Model regresji możemy zapisać następująco

$$Y = \beta X + \varepsilon.$$

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),]
```

Macierz x wygląda następująco

$$X = \begin{bmatrix} 1 & 19 \\ 1 & 10 \\ 1 & 7 \\ 1 & 24 \\ 1 & 14 \\ 1 & 18 \\ 1 & 15 \\ 1 & 7 \\ 1 & 22 \\ 1 & 18 \\ 1 & 11 \\ 1 & 20 \\ 1 & 13 \\ 1 & 17 \\ 1 & 25 \\ 1 & 12 \\ 1 & 24 \\ 1 & 20 \\ 1 & 12 \\ 1 & 15 \end{bmatrix}$$

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

$$(X^T X)^{-1} X^T Y = \left(\begin{bmatrix} 1 & 19 \\ 1 & 10 \\ 1 & 7 \\ 1 & 24 \\ 1 & 14 \\ 1 & 18 \\ 1 & 15 \\ 1 & 7 \\ 1 & 22 \\ 1 & 18 \\ 1 & 11 \\ 1 & 20 \\ 1 & 13 \\ 1 & 17 \\ 1 & 25 \\ 1 & 12 \\ 1 & 24 \\ 1 & 20 \\ 1 & 12 \\ 1 & 15 \end{bmatrix} \begin{bmatrix} 1 & 19 \\ 1 & 10 \\ 1 & 7 \\ 1 & 24 \\ 1 & 14 \\ 1 & 18 \\ 1 & 15 \\ 1 & 7 \\ 1 & 22 \\ 1 & 18 \\ 1 & 11 \\ 1 & 20 \\ 1 & 13 \\ 1 & 17 \\ 1 & 25 \\ 1 & 12 \\ 1 & 24 \\ 1 & 20 \\ 1 & 12 \\ 1 & 15 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 19 \\ 1 & 10 \\ 1 & 7 \\ 1 & 24 \\ 1 & 14 \\ 1 & 18 \\ 1 & 15 \\ 1 & 7 \\ 1 & 22 \\ 1 & 18 \\ 1 & 11 \\ 1 & 20 \\ 1 & 13 \\ 1 & 17 \\ 1 & 25 \\ 1 & 12 \\ 1 & 24 \\ 1 & 20 \\ 1 & 12 \\ 1 & 15 \end{bmatrix}^T \begin{bmatrix} 46 \\ 18 \\ 22 \\ 92 \\ 36 \\ 42 \\ 54 \\ 4 \\ 66 \\ 76 \\ 28 \\ 64 \\ 26 \\ 40 \\ 85 \\ 24 \\ 93 \\ 32 \\ 20 \\ 26 \end{bmatrix}$$

Co daje nam

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \left(\begin{bmatrix} 20 & 323 \\ 323 & 5781 \end{bmatrix} \right)^{-1} \begin{bmatrix} 894 \\ 16855 \end{bmatrix} = \begin{bmatrix} -24.439907891242 \\ 4.28110884775491 \end{bmatrix}$$

Rysujemy wykres regresji

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

Zależność między dystansem, a prędkością

