Tutorial 6

AID-521: Mathematics for Data Science

Abhishek Samantray, Vimal Kumar

September 6, 2021

Qn. 1.

The following data relate to the prices (Y) of five randomly chosen houses in a certain neighborhood, the corresponding ages of the houses (x_1), and square footage (x_2).

Price y (in 1000 \$)	Age x_1 (in years)	Square footage x_2 (in 1000s of sq. ft.)
100	1	1
80	5	1
104	5	2
94	10	2
130	20	3

(a) Fit a multiple linear regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

to the above data points.

Ans/Sol.

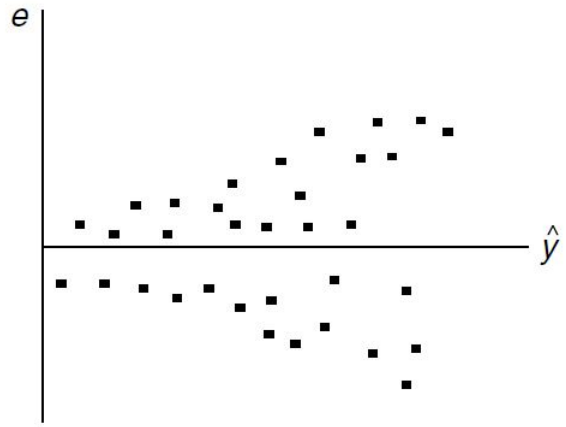
Qn. 2.

A regression was specified on a set of data points $i = 1, ..., n$ containing independent variable Y and a dependent variable X_1 as

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

where the error term ϵ has the usual assumptions of $E[\epsilon] = 0$ and $Var(\epsilon) = \sigma^2$ (a constant).

(a) The model was estimated using least squares. Post-estimation, the following graph shows the co-variation of the estimated error terms e_i and the predicted values $\hat{y}_i$.



Testing for the significance of the estimated value of $\hat{\beta}_1 = 1.5$, the slope parameter, it was found to be statistically significant at $\alpha = 0.05$. Is $\hat{\beta}_1$, the estimator of β_1, unbiased? Can you rely on the estimated value of β, and the significance of $\hat{\beta}_1$?

Ans/Sol.