Statement:

Prove that the covariance between residuals and predictor variable is zero for a linear regression model.

Restatement:

Restating the problem, given $(x_1, y_1),(x_2\ y_2),.....(x_{n-1}, y_{n-1}),(x_n, y_n)$, the linear regression model is given by

$$y = a_0 + a_1 x$$

where

$$a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^{\,2} - (\sum_{i=1}^{n} x_i)^2} \qquad (1)$$

$$a_0 = \bar{y} - a_1\bar{x} \qquad (2)$$

where

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad (3)$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (4)$$

If the residual at a point x_i is given by

$$E_i = y_i - a_0 - a_1 x_i \qquad (5)$$

Then show that

$$\sum_{i=1}^{n} E_i x_i = 0 \qquad (6)$$

Using Equation (5) gives

$$\sum_{i=1}^{n} E_i x_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)(x_i)$$

Using Equation (2) gives

$$\sum_{i=1}^{n} E_i x_i = \sum_{i=1}^{n} (y_i - (\bar{y} - a_1\bar{x}) - a_1 x_i)(x_i)$$

$$= \sum_{i=1}^{n} (y_i - \bar{y} + a_1 \bar{x} - a_1 x_i)(x_i)$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})(x_i) + a_1 \sum_{i=1}^{n} (\bar{x} - x_i)(x_i)$$

Using Equations (3) and (4) for $\bar{x}$ and $\bar{y}$ gives

$$\sum_{i=1}^{n} E_i x_i = \sum_{i=1}^{n} \left(y_i - \frac{\sum_{i=1}^{n} y_i}{n} \right)(x_i) + a_1 \sum_{i=1}^{n} \left(\frac{\sum_{i=1}^{n} x_i}{n} - x_i \right)(x_i)$$

$$= \sum_{i=1}^{n} x_i\, y_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n} + a_1 \left(\frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n} - \sum_{i=1}^{n} x_i^{\,2} \right)$$

$$= \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n} + a_1 \frac{\left(\sum_{i=1}^{n} x_i \right)^2 - n\sum_{i=1}^{n} x_i^{\,2}}{n}$$

Using Equation (1) for a_1 gives

$$\sum_{i=1}^{n} E_i x_i = \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n}$$

$$+ \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i \right)^2} \times \frac{\left(\sum_{i=1}^{n} x_i \right)^2 - n\sum_{i=1}^{n} x_i^{\,2}}{n}$$

$$= \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n} - \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}$$

$$= \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i - n\sum_{i=1}^{n} x_i\, y_i + \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}$$

$$= 0$$