In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1,2,...,n$$
$$E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2 \quad and \quad Cov(\varepsilon_i, \varepsilon_j) = 0 \ for\ all\ i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad and \quad Var(Y_i) = \sigma^2.$$

Let's introduce some more notations:

$$S_{xx} = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$
$$S_{yy} = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$
$$S_{xy} = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$$

The least square estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{S_{xy}}{S_{xx}}$$
$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1\bar{X}$$

Properties of Fitted Regression line

The residuals $e_i = Y_i - \hat{Y}_i,\ i = 1,2,...$ of the simple linear regression model satisfies the following properties

1- The sum of the residuals is zero i.e.

$$\sum_{i=1}^{n} e_i = 0$$

Proof.

$$\sum e_i = \sum \left(Y_i - \widehat{Y}_i \right) = \sum \left(Y_i - (b_0 + b_1 X_i) \right)$$

$$= \sum \left(Y_i - b_0 - b_1 X_i \right)$$

$$= n\bar{Y} - nb_0 - nb_1\bar{X}$$
$$= n\bar{Y} - n(\bar{Y} - b_1\bar{X}) - nb_1\bar{X}$$
$$= n\bar{Y} - n\bar{Y} + nb_1\bar{X} - nb_1\bar{X}$$
$$= 0$$

2- The regression line always goes through the point $\left(\bar{X}, \bar{Y}\right)$.

Proof.

$$Y_i = b_0 + b_1 X_i = b_0 + b_1\bar{X} = \bar{Y} - b_1\bar{X} + b_1\bar{X} = \bar{Y}$$

3- The sum of the observed values Y_i ; equals the sum of the fitted values $\widehat{Y}_i$

Proof.

$$\sum \widehat{Y}_i = \sum (b_0 + b_1 X_i) = nb_0 + nb_1\bar{X}$$

$$= n(\bar{Y} - b_1\bar{X}) + nb_1\bar{X} = n\bar{Y} - nb_1\bar{X} + nb_1\bar{X}$$

$$= n\bar{Y} = \sum Y_i$$

4- The sum of the weighted residuals is zero when the residual in the ith trial is weighted by the level of the predictor variable in the ith trial $\sum_{i=1}^{n} e_i X_i = 0$

Proof.

$$\begin{aligned}
\sum_{i=1}^{n} e_i X_i &= \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right) X_i = \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_i\right) X_i \\
&= \sum_{i=1}^{n} \left(Y_i - (\bar{Y} - b_1 \bar{X}) - b_1 X_i\right) X_i \\
&= \sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i (X_i - \bar{X}) \\
&= \sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i - b_1 \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \\
&= \sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right) \\
&= \sum_{i=1}^{n} Y_i X_i - n\bar{X}\bar{Y} - \sum_{i=1}^{n} Y_i X_i + n\bar{X}\bar{Y} \\
&= 0.
\end{aligned}$$

5- A consequence of properties (1) and (4) is that the sum of the weighted residuals is zero when the residual in the ith trial is weighted by the fitted value of the response variable for the ith trial $\sum_{i=1}^{n} e_i \hat{Y}_i$.

Proof.

$$\sum_{i=1}^{n} e_i \hat{Y}_i = \sum_{i=1}^{n} e_i (b_0 + b_1 X_i) = b_0 \sum_{i=1}^{n} e_i + b_1 \sum_{i=1}^{n} e_i X_i = 0 + 0 = 0$$

Point Estimator of σ^2

Under the assumption that the residuals $\varepsilon_i \sim N(0, \sigma^2)$, the maximum likelihood methods can be used to derive the MLE of σ^2 as

$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n} = \frac{SSE}{n}$ which is biased estimate for σ^2

The unbiased estimate of σ^2 can be obtained as

$S^2 = MSE = \frac{n}{n-2}\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2} = \frac{SSE}{n-2}$ which is called the residual mean square MSE

Then

$$\boxed{MSE = \frac{SSE}{n-2}}$$

Example

We will calculate *SSE* for the Toluca Company example. The residuals were obtained earlier. From these results, we obtain:

$$\boldsymbol{SSE = 54825}$$

Then

$$MSE = S^2 = \frac{54825}{25\text{-}2} = 2384 \quad \text{and} \quad S = \sqrt{MSE} = \sqrt{2384} = 48.8 \text{ hours}$$

The sum of squared errors can be calculated from R results simply as:

```
model=lm(y~x)
summary(model)
summary(model)$coef
e=summary(model)$res
sum(e^2)
```