Residual Sum of Squares

- Its the sum of rest of $e_i = y_i - \hat{y}_i$

$$RSS = e_1^2 + e_2^2 + \cdots + e_n^2 = \sum_{i=1}^{n} e_i^2$$

- The RSS expression is derived from the Maximum Likelihood Estimation by assuming that the error terms takes on a normal distribution bell curve

$$\hat{\beta}^{\text{MLE}} = \arg\max_{\beta} \sum_{i=1}^{n} \log[p(X_i|\beta)]$$

RSS Derivation

1. Assumption of **Linear Regression**

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$
- i.i.d.(independent,identically distributed)

1. **Likelihood Function**

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}$$

Where :

- $x_i = Y_i$ The observed Data
- $\mu = \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ The expected value
- The normal distribution is on Y

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(Y_i-\beta_0-\beta_1 X_i)^2/2\sigma^2}$$

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\sum(y_i-\hat{y}_i)^2/2\sigma^2}$$

- $(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$$\begin{aligned}
&\text{Log Likelihood} \\
&\log(L(\beta, \sigma^2)) = \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} - \sum (y_i - \hat{y}_i)^2/2\sigma^2 \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2/2\sigma^2
\end{aligned}$$

$$\begin{aligned}
&\text{Maximazing with respect to } \beta_0 \\
&\frac{dL}{d\hat{\beta}_0} = \frac{1}{\sigma^2} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) \\
&\frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 X_i) - \hat{\beta}_0 \\
&\text{Soliving For 0} \\
&\frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 X_i) - n\hat{\beta}_0 = 0 \\
&n\hat{\beta}_0 = \sum y_i - \hat{\beta}_1 X_i = n\bar{y} - \hat{\beta}_1 n\bar{x} \\
&\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}
\end{aligned}$$

$$
\text{Maximazing with respect to } \beta_1
$$

$$
\begin{aligned}
\frac{dL}{d\hat{\beta}_1} &= -\frac{1}{2\sigma^2}\sum(-2X_i(y_i-\hat{\beta}_0-\hat{\beta}_1X_i)) \\
\frac{dL}{d\hat{\beta}_1} &= \frac{1}{\sigma^2}\sum(X_i(y_i-\bar{y}+\hat{\beta}_1\bar{x}-\hat{\beta}_1X_i))
\end{aligned}
$$

$$
\text{Soliving For 0}
$$

$$
\begin{aligned}
&\frac{1}{\sigma^2}\sum(X_i(y_i-\bar{y}+\hat{\beta}_1\bar{x}-\hat{\beta}_1X_i)) = 0 \\
&\sum X_i(y_i-\hat{y}) - \hat{\beta}_1\sum Xi(x_i-\bar{x}) = 0 \\
&\hat{\beta}_1 = \frac{\sum y_i-\hat{y}}{\sum x_i-\bar{x}} \\
&\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \\
&\hat{\beta}_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var(X)}}
\end{aligned}
$$

- $f^n\ \frac{d}{df^n} = n\frac{d}{df}f^{n-1}$
- When maximizing a **Likelihood function** and solving for 0 we ignore the constant terms $\frac{1}{\sigma^2}$

$$
\text{Maximazing with respect to } \sigma^2
$$

$$
\frac{dL}{d\hat{y}_i} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum(y_i-\hat{y}_i)^2
$$

$$
\text{Soliving For 0}
$$

$$
\begin{aligned}
&-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum(y_i-\hat{y}_i)^2 = 0 \\
&\frac{n}{2\sigma^2} = \frac{1}{2(\sigma^2)^2}\sum(y_i-\hat{y}_i)^2 \\
&\frac{2(\sigma^2)^2}{2\sigma^2} = \frac{\sum_{i=1}^n(y_i-\hat{y}_i)^2}{n} \\
&\sigma^2 = \frac{\sum_{i=1}^n(y_i-\hat{y}_i)^2}{n} = \frac{\mathrm{RSS}}{n} = \mathrm{MSE}
\end{aligned}
$$