



종속변수의 편회분산. (=총분산)

우선 SST를 먼저 설명하고자 한다. SST는 Sum of Square Total로 편차의 제곱합으로, $SST = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$ 로 표현할 수 있다. 쉽게 이야기하면 모든 차이를 제공하여 더한 것이다. 무슨^①차이를 더한 것일까? 실제값과 예측값, 평균값 사이에 발생한 차이이며, 이것이 SSE와 SSR이다.

- ① SSE는 Sum of Square Error로, 회귀식과 실제값의 차이를 의미한다. 위의 그림을 다시 확인해보면, 파란선 \hat{y} 와 검은점 사이에 유격이 존재하는 것을 알 수 있을 것이다. 이 유격을 제공해 더한 것이 SSE다. 식으로는 $SSE = \sum (y_i - \hat{y}_i)^2$ 로 나타낼 수 있다. 이 유격이 작을수록 \hat{y} 이 모든 데이터를 고르게 설명한다고 해석할 수 있고, 높은 R-squared 값을 도출할 수 있다.
- ② SSR은 Sum of Square Regression으로, 회귀식과 평균값의 차이를 의미한다. 위의 그림에서 파란선 \hat{y} 과 빨간선 \bar{y} 이 엇갈린 차이가 발생하는 것을 확인할 수 있다. 이를 식으로 나타내면 $SSR = \sum (\hat{y}_i - \bar{y})^2$ 이다. 평균, 즉 \bar{y} 와 차이가 날수록 SSR의 값이 커지는데, 이는 \hat{y} 가 모든 데이터를 고루 설명하고 있다는 것으로 해석할 수 있다. SSR이 높아질수록 R-squared가 높아지는 것이다.

자, SST는 모든 차이의 합이며, 이것이 SSE와 SSR이라고 하였다. SST의 수식과 SSE, SSR의 수식을 비교해보길 바란다. 이미 SST 안에 SSE와 SSR이 들어있는 것을 확인할 수 있을 것이다. 즉, $SST = SSE + SSR$ 로도 표현할 수 있다.

• R^2 (결정계수): 회귀식이 얼마나 종속변수를 잘 설명하고 있는지를 나타내는 것.

$$R^2 = \frac{SSR}{SST}$$

$$\text{TSS} = \text{SST} := \sum_{i=1} (y_i - \bar{y})^2$$

7. ESS(Explained Sum of Squares) 혹은
SSR(Sum of Squares due to Regression) :

$$\text{ESS} = \text{SSR} := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

8. RSS(Residual Sum of Squares) 혹은
SSE(Sum of squared Error) :

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

9. 중회귀계수 R-squared 혹은 설명력 :

$$R^2 := \frac{\text{SSR}}{\text{SST}}$$

✂

~~1~~ 1 = total variability = Total Sum Sq

2 = variability (attributed to class) = between group variability = explained variability = class Sum Sq

3 = variability (attributed to other factors) = within group variability = Residuals Sum Sq

What is Explained Sum of Square (ESS)?

Explained sum of square (ESS) or **Regression sum of squares** or Model sum of squares is a statistical quantity used in modeling of a process. ESS gives an estimate of how well a model explains the observed data for the process.

It tells how much of the variation between observed data and predicted data is being explained by the model proposed.

Mathematically, it is the sum of the squares of the difference between the predicted data and mean data.

$$\rightarrow \bar{y}$$

What is Explained Sum of Square (ESS)?

Explained sum of square (ESS) or **Regression sum of squares** or Model sum of squares is a statistical quantity used in modeling of a process. ESS gives an estimate of how well a model explains the observed data for the process.

It tells how much of the variation between observed data and predicted data is being explained by the model proposed. Mathematically, it is the sum of the squares of the difference between the predicted data and mean data.

Let $y_i = a + b_1x_{1i} + b_2x_{2i} + \dots + \varepsilon_i$ is regression model, where:

y_i is the i^{th} observation of the response variable

x_{ji} is the i^{th} observation of the j^{th} explanatory variable

a and b_j are coefficients

i indexes the observations from 1 to n

ε_i is the i^{th} value of the error term

Then

$$\text{ESS} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

This is usually used for regression models. The variation (in the modeled values) is contrasted with the variation (in the observed data (total sum of squares)) and variation in modeling errors (residual sum of squares). The result of this comparison is given by ESS as per the following equation:

↑ $R^2 = \frac{ESS}{SST}$, 만약 SSR이 SST와 매우 비슷하게 된다면 해당 model의 설명력은 높은 것이다.

ESS = total sum of squares – residual sum of squares

As a generalization, a high ESS value signifies greater amount of variation being explained by the model, hence meaning a better model.