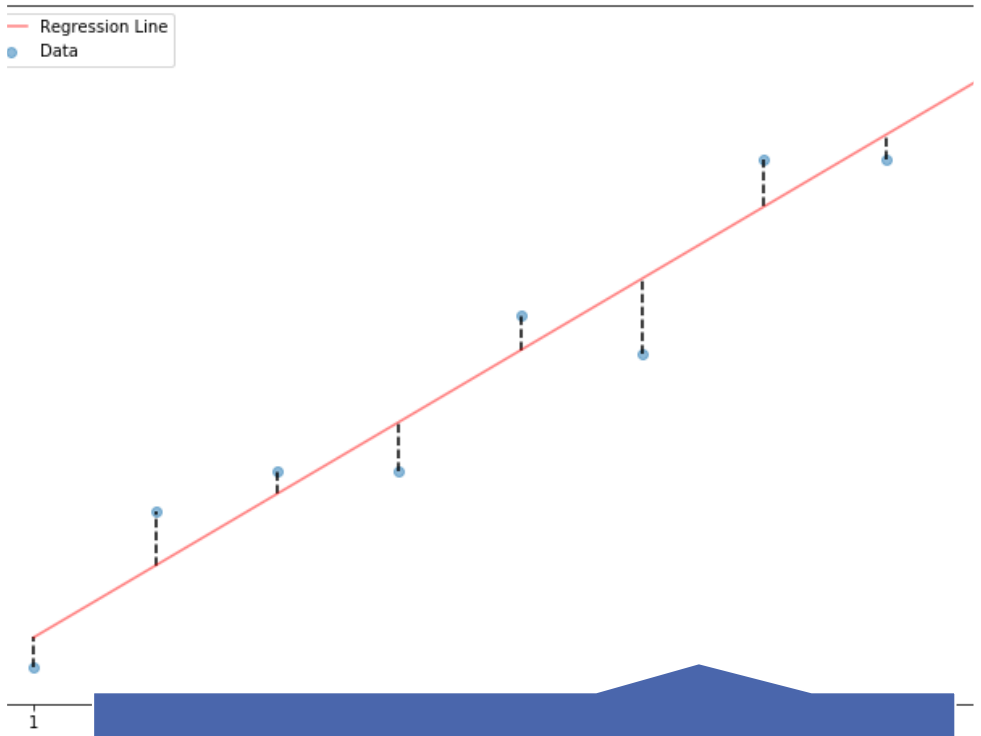
Regression – Analysis of Variance (ANOVA)



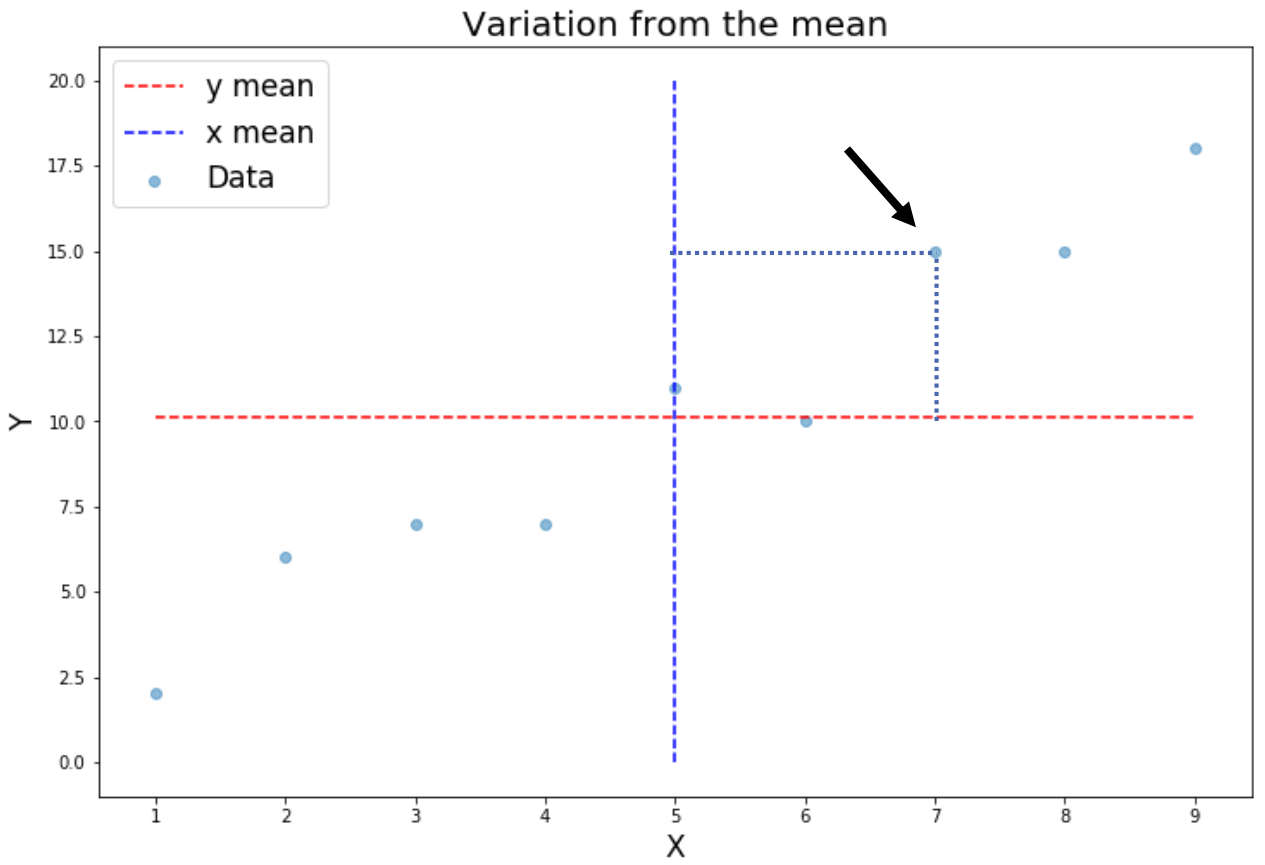
How well does the model fit the data & explain the dependent variable?

Analysis of Variance (ANOVA)

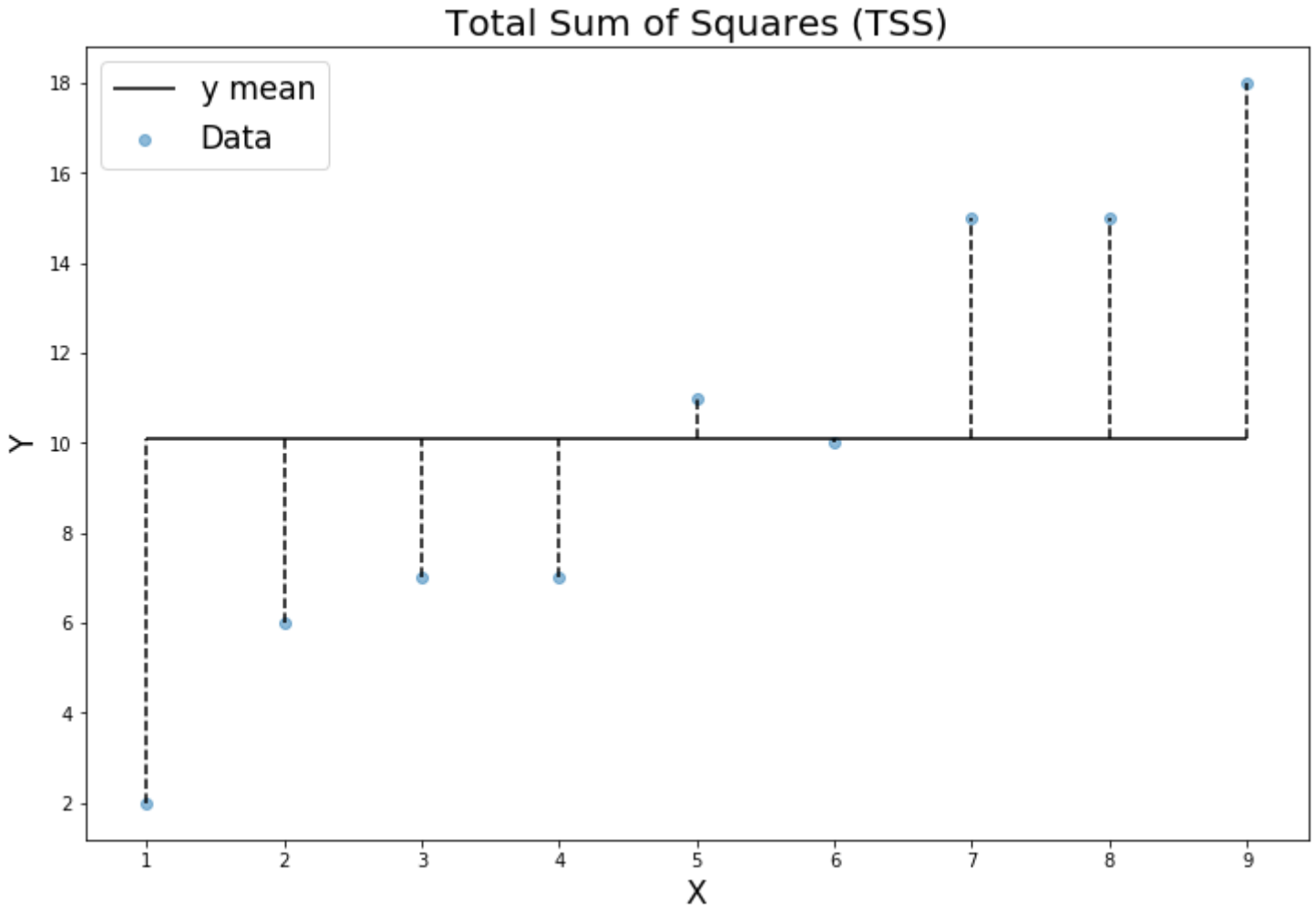
Linear Regression – Variation

The purpose of simple linear Regression is to explain the variation in a dependent variable by the variation in a single independent variable.

"Variation" is interpreted as degree to which a variable differs from its mean value (similar but not the same as variance).



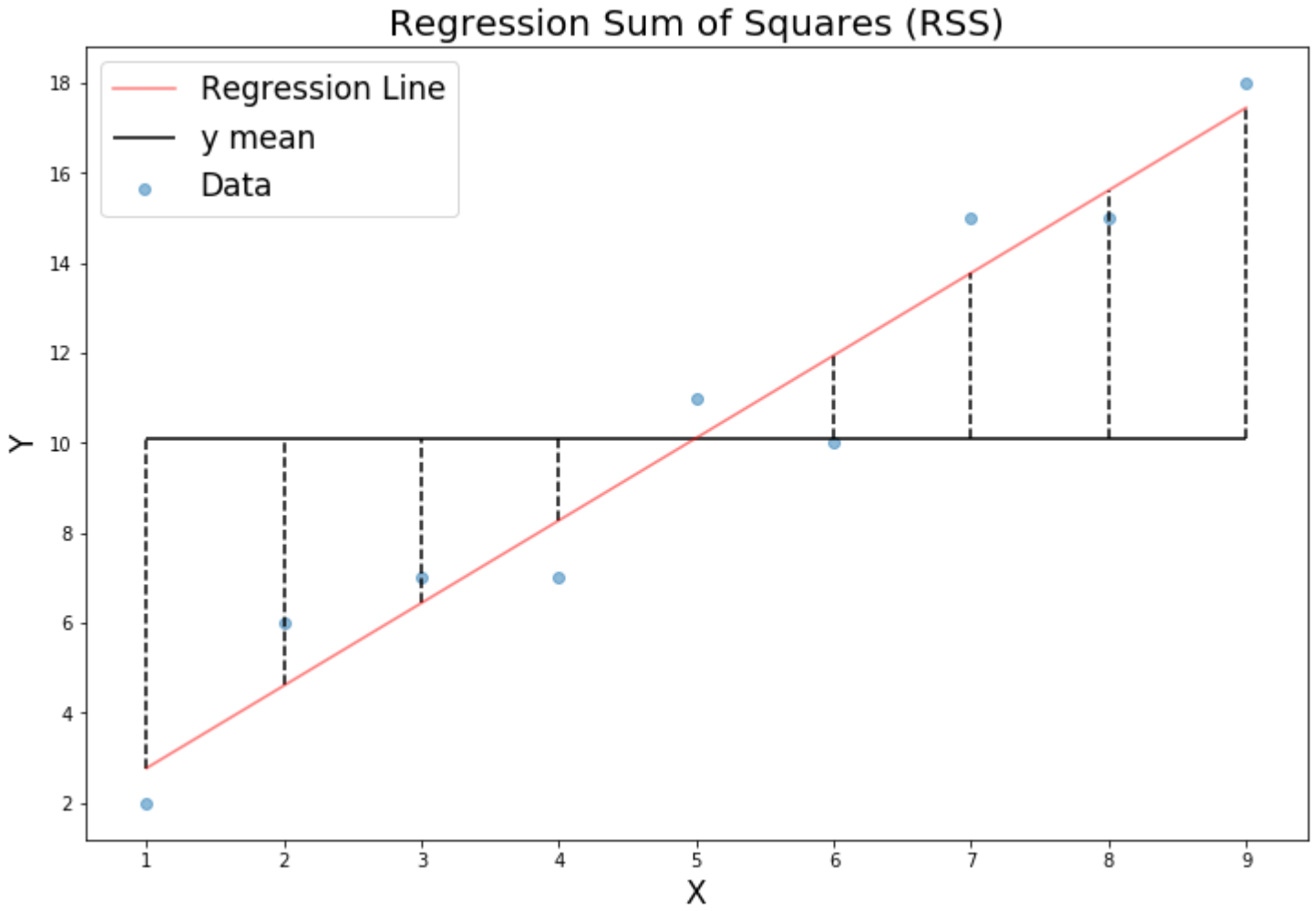
Total Variation in dependent variable (TSS)



$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 212.9$$

n: sample size
Y_i: ith observation on variable Y
$\bar{Y}$: mean of the variable Y

Variation in dependent variable explained by model (RSS)



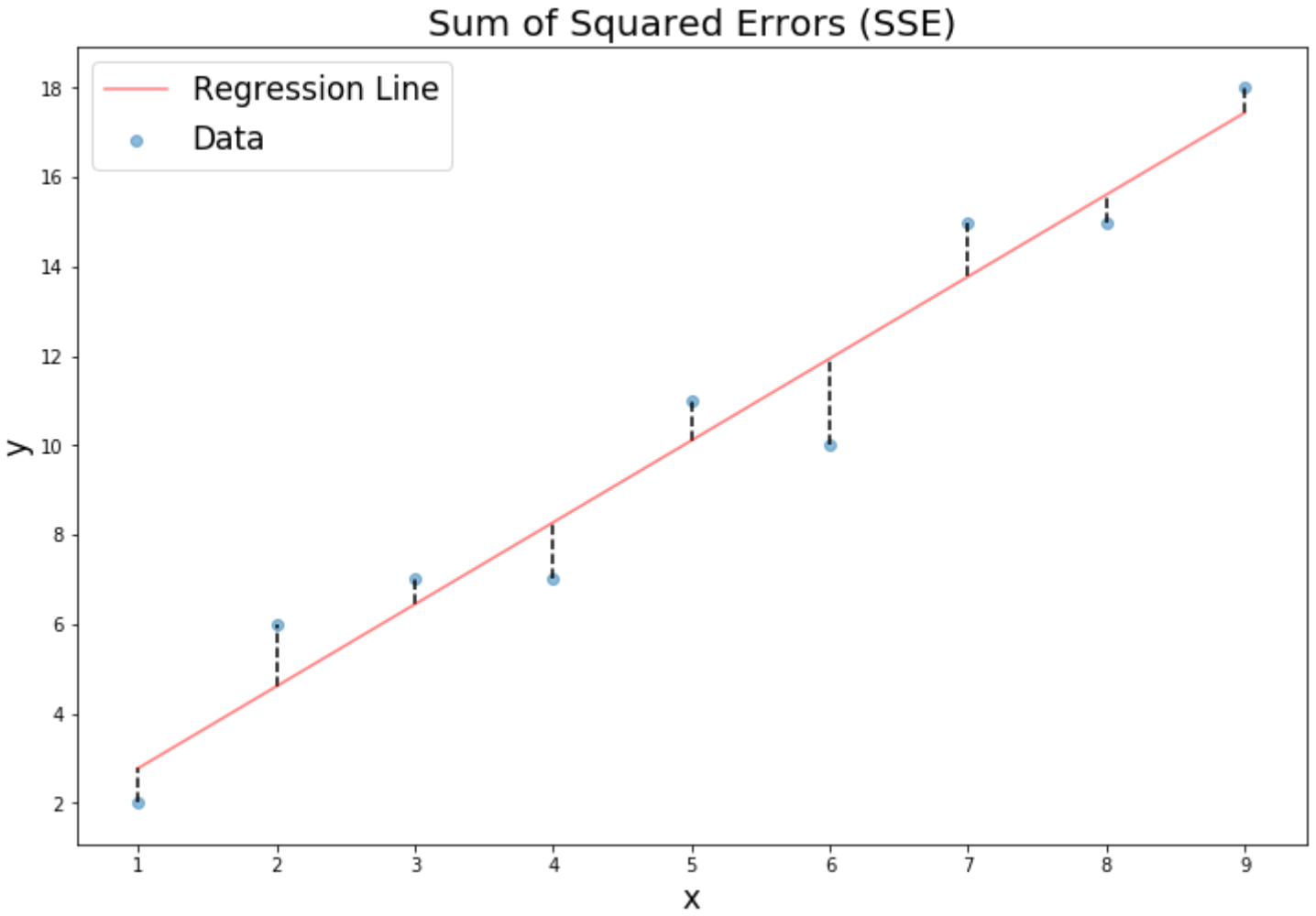
$$RSS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = 201.7$$

n: sample size

$\hat{Y}_i$: estimated model value of Y_i

$\bar{Y}$: mean of the variable Y

Unexplained Variation in dependent variable (SSE)



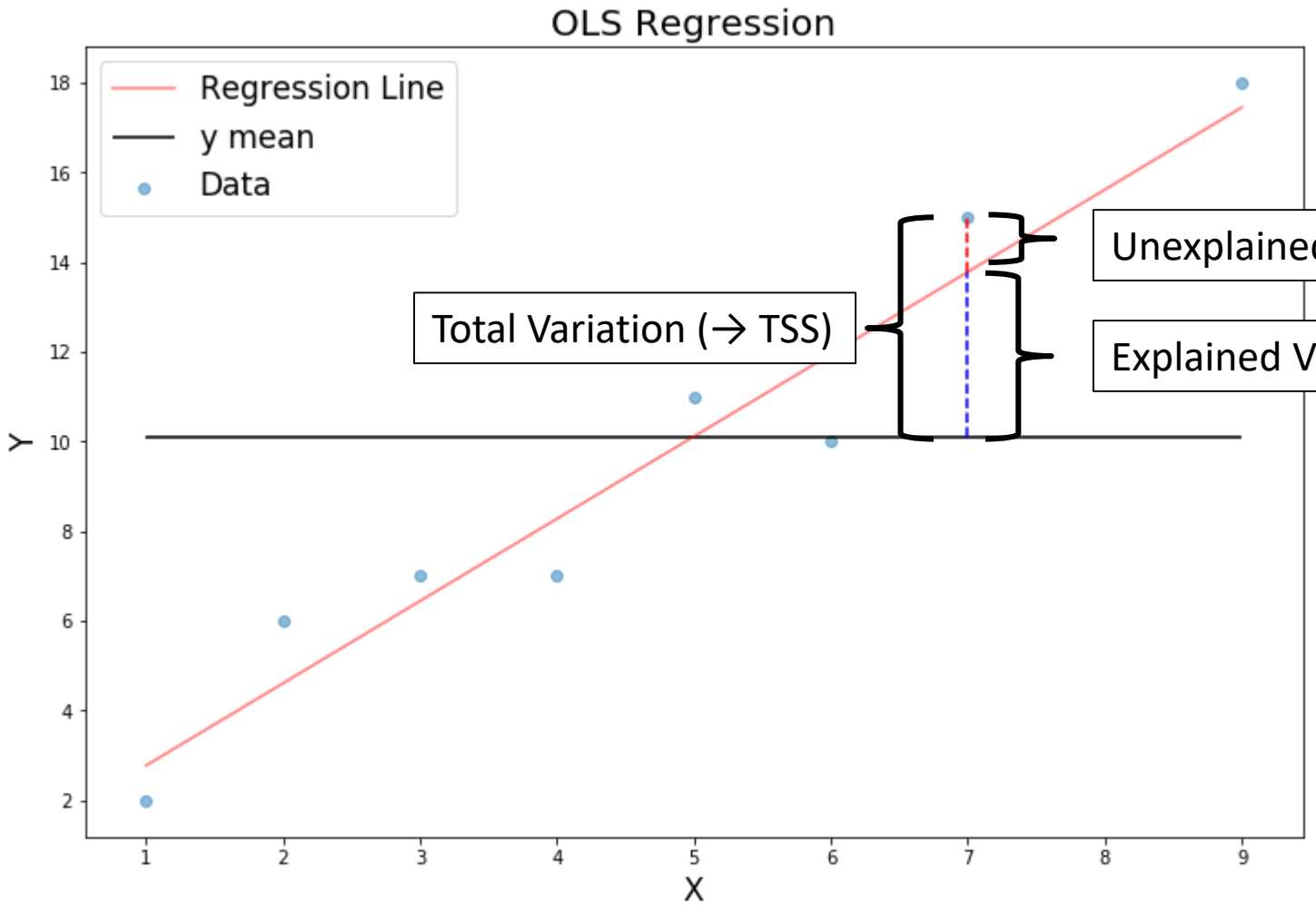
$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = 11.2$$

n: sample size

Y_i: ith observation on variable Y

$\hat{Y}_i$: ith observation on variable Y

ANOVA - Summary



Min!

$$SSE = TSS - RSS$$
$$11.2 = 212.9 - 201.7$$

Coefficient of Determination - R^2(R squared)

R^2 is defined as the proportion of the variation in the dependent variable that is explained by the independent variable / regression model. It gives some information about the goodness of fit of a model and can range from 0 (no fit) to 1 (perfect fit).

$$R^2 = \frac{Total\ Variation\ (TSS) - Unexplained\ Variation\ (SSE)}{Total\ Variation}$$

$$= \frac{Explained\ Variation\ (RSS)}{Total\ Variation\ (TSS)} = \frac{201.7}{212.9} = 0.947$$

With only one independent Variable:

$$R^2 = r^2 = 0.973^2 = 0.947$$

How well does the model fit the data & explain the dependent variable?