Regression: Simple Version

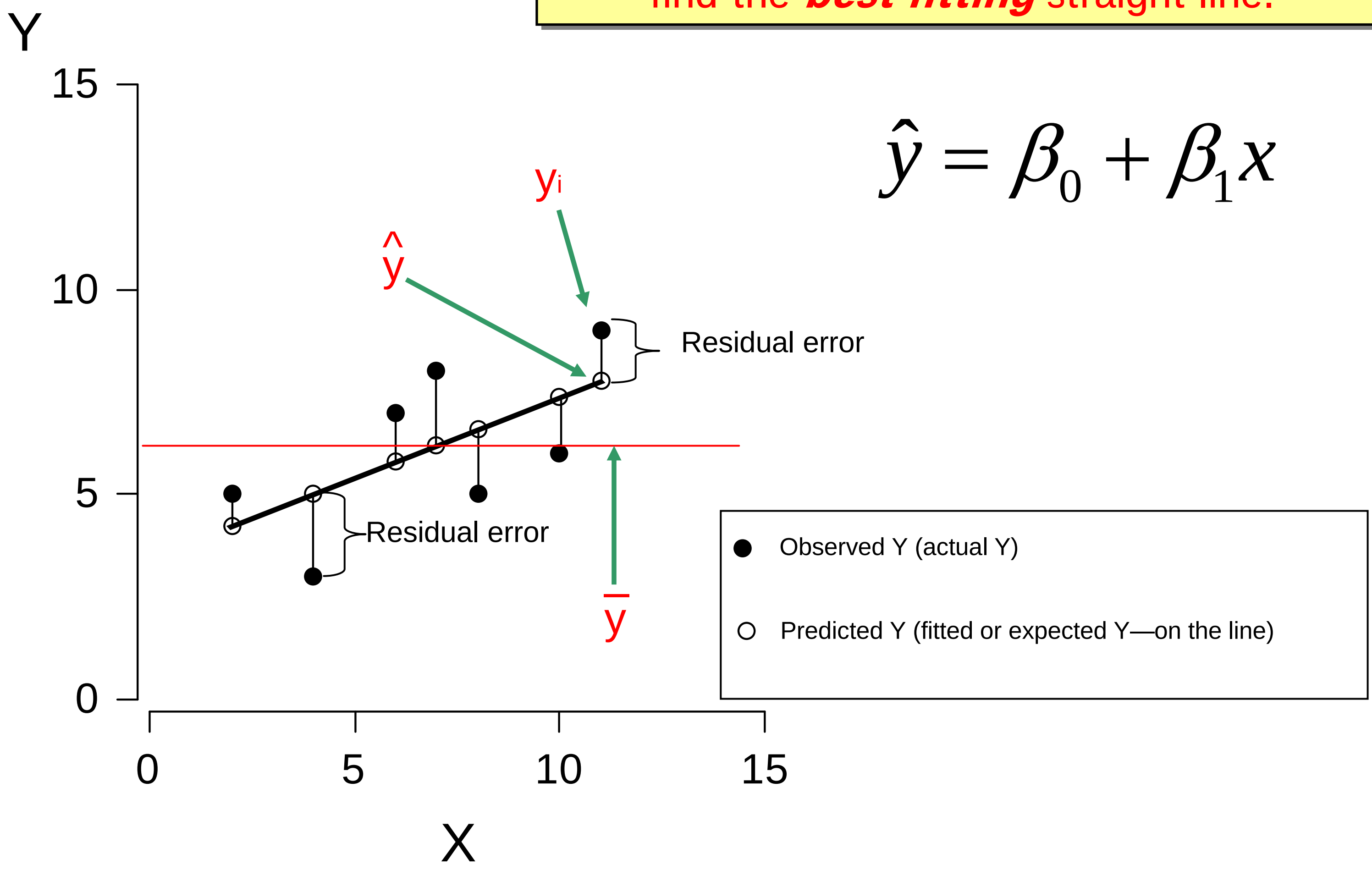
The Lean Six Sigma Institute

www.lean-six-sigma.pro

REGRESSION LINE

We use the **method of least squares** to find the ***best fitting*** straight line.

$$\hat{y} = \beta_0 + \beta_1 x$$

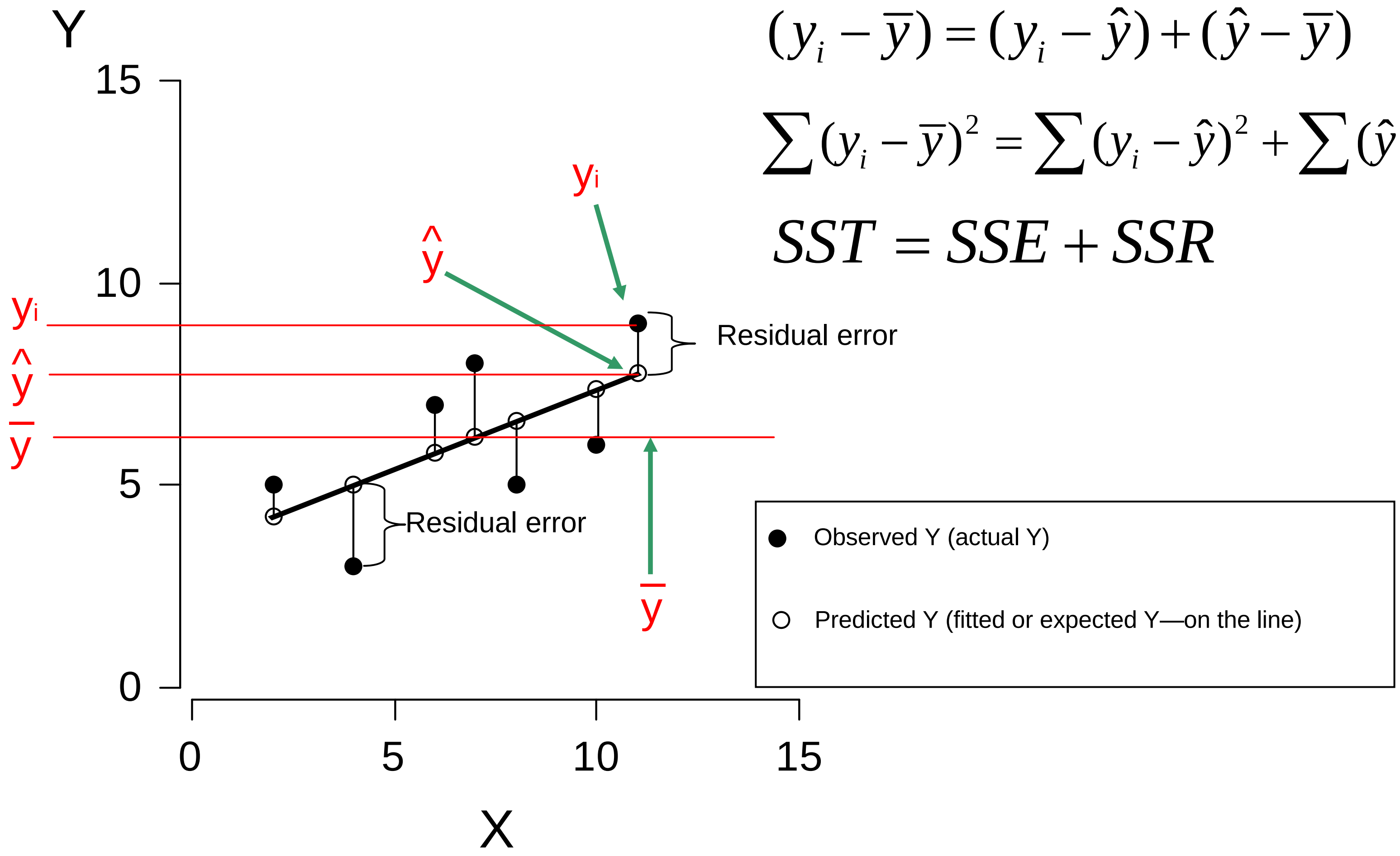


REGRESSION SUMS OF THE SQUARES

$$(y_i - \bar{y}) = (y_i - \hat{y}) + (\hat{y} - \bar{y})$$

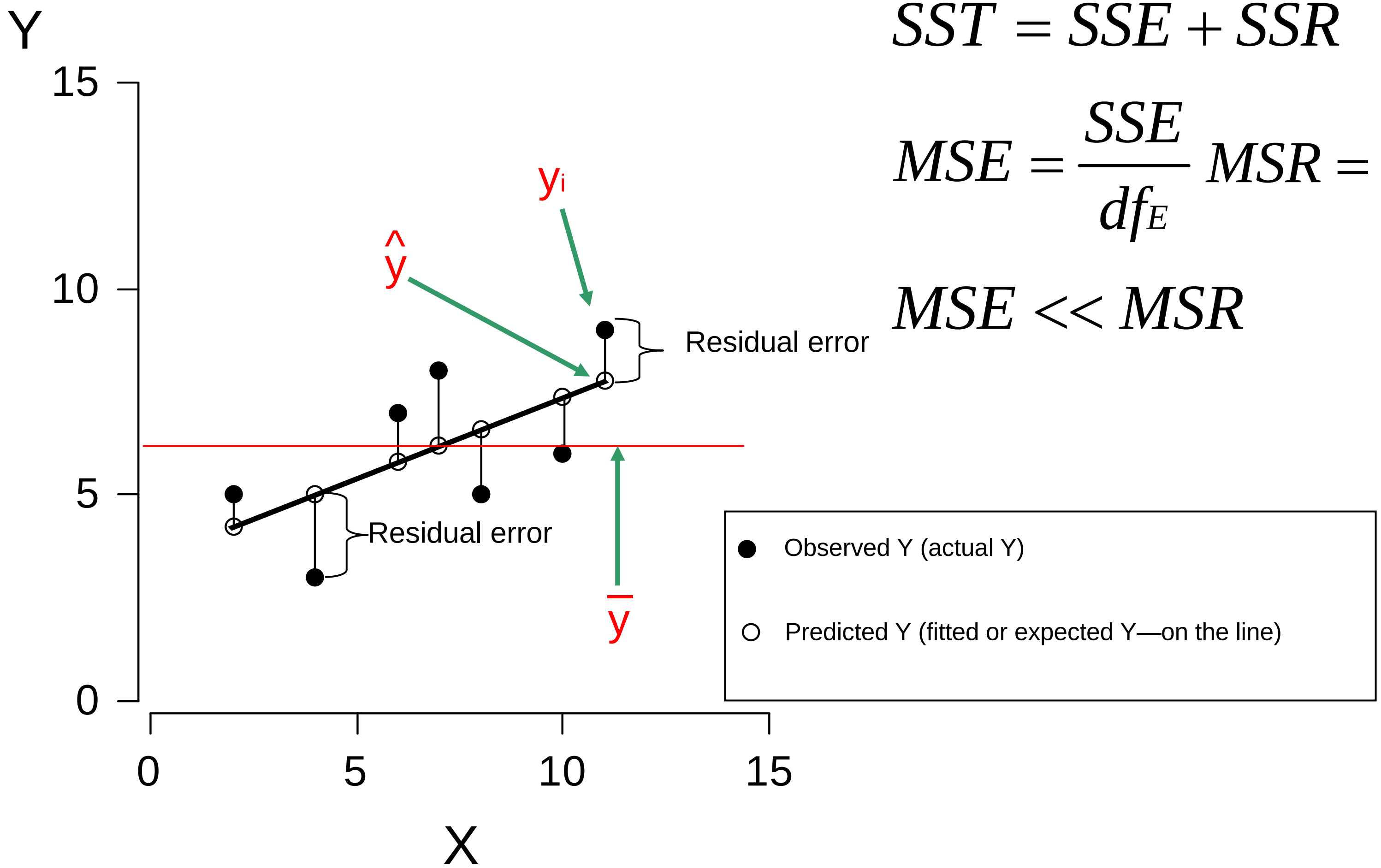
$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y})^2 + \sum (\hat{y} - \bar{y})^2$$

$$SST = SSE + SSR$$



We find the ***regression sums of squares***.

MEAN SQUARE ERROR AND MEAN SQUARE REGRESSION



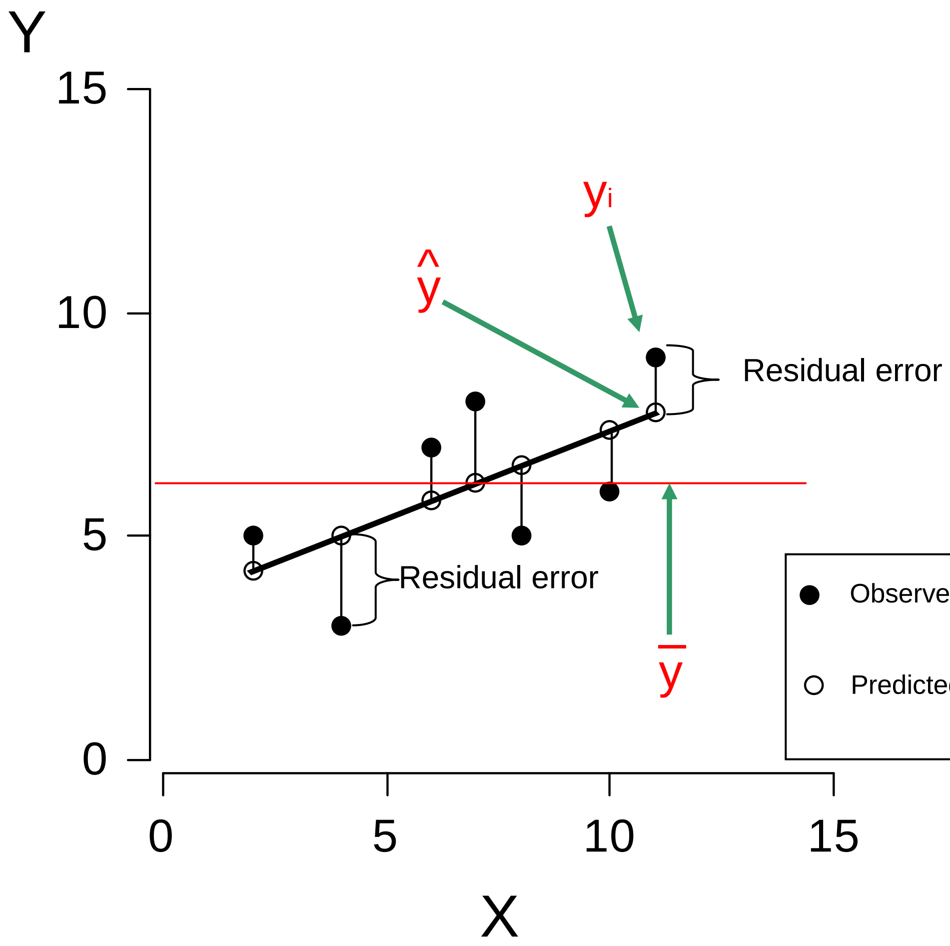
$$SST = SSE + SSR$$

$$MSE = \frac{SSE}{df_E} \quad MSR = \frac{SSR}{df_R}$$

$$MSE << MSR$$

To find the ***best fitting*** straight line, the variance of the error (**Mean Square Error)** about the Regression Line must be **significantly less than the variance of y** about the mean.

REGRESSION AND R^2



$$SST = SSE + SSR$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

REGRESSION	R^2	SSE =
NONE	0	SST
PERFECT	1	0

● Observed Y (actual Y)

○ Predicted Y (fitted or expected Y—on the line)

We use R^2 to **check how much of the variability is explained by the regression line**.

REGRESSION F STATISTIC AND HYPOTHESIS TEST

We then use the F statistic with the ANOVA table to find the p value and test the Null Hypothesis

$$\hat{y} = \beta_0 + \beta_1 x$$

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

$$SST = SSE + SSR$$

$$MSR = \frac{SSR}{df_R}$$

$$MSE = \frac{SSE}{df_E}$$

$$s_y^2 = \frac{SST}{df_T}$$

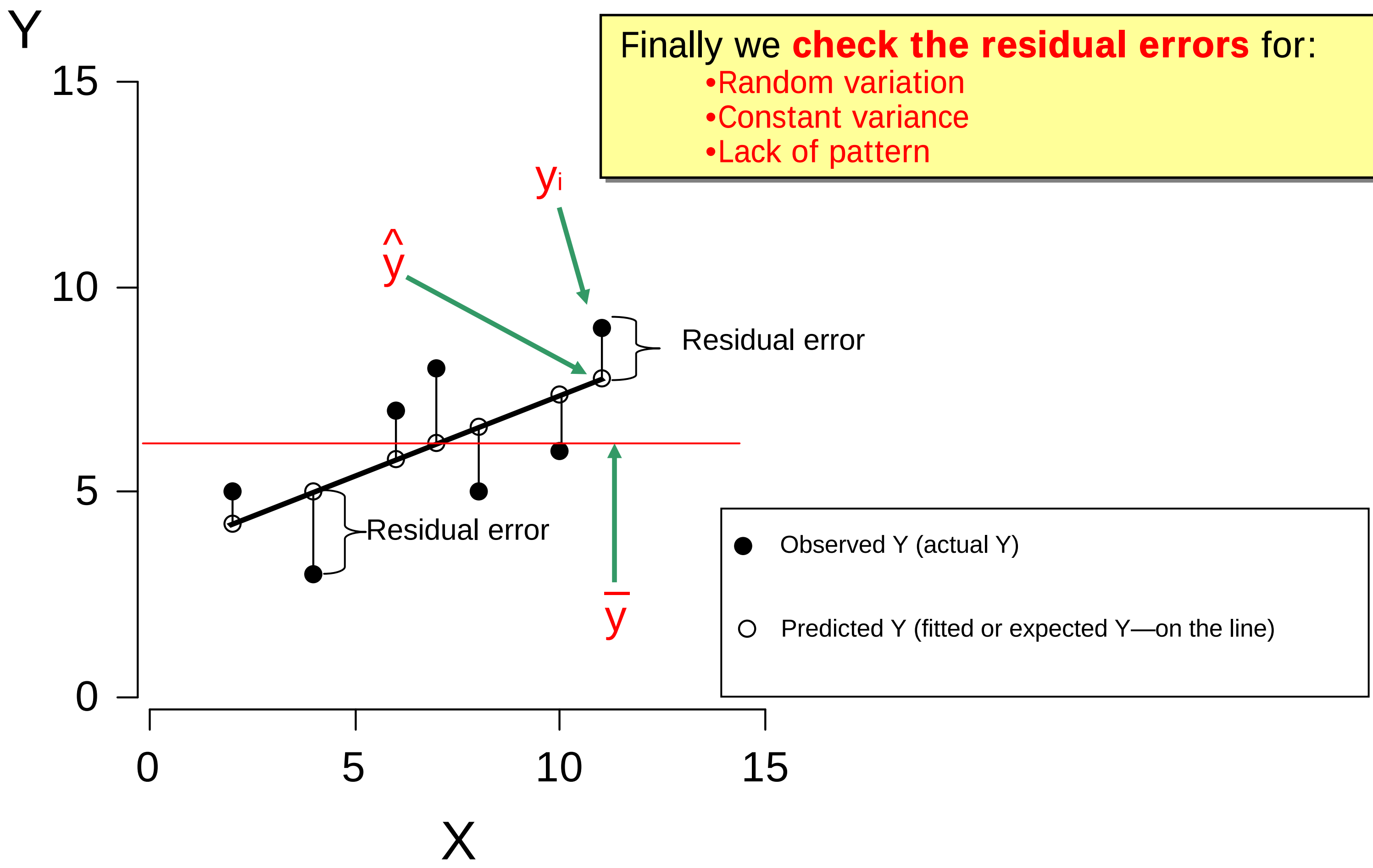
	Sum of Squares	df	Mean Square	F value
Regression	SSR	df_R	MSR	MSR/MSE
Error	SSE	df_E	MSE	
Total	SST	df_T		

$$MSE << MSR$$

REGRESSION

Finally we **check the residual errors** for:

- Random variation
- Constant variance
- Lack of pattern

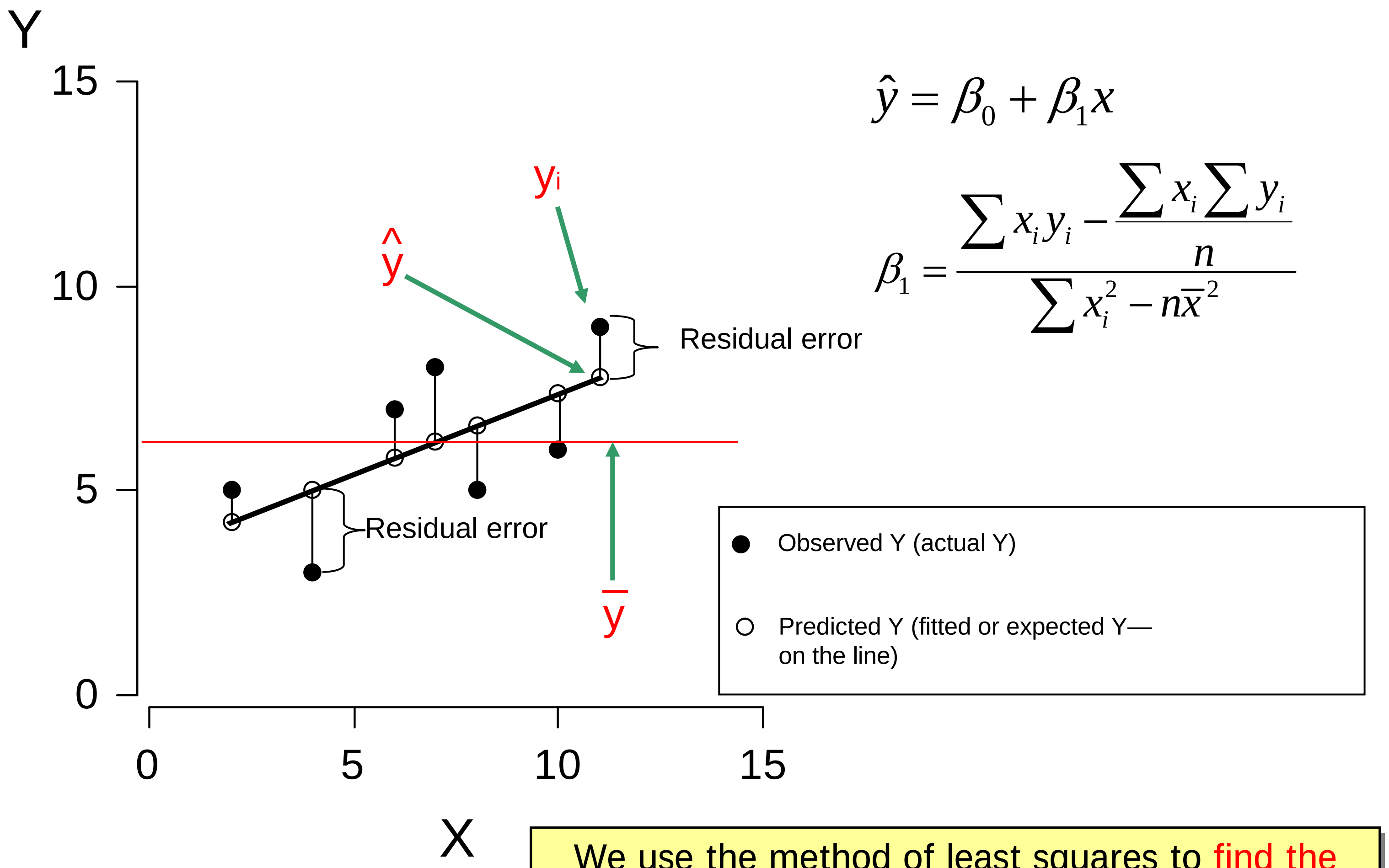


Regression: Version citing statistical formulae

The Lean Six Sigma Institute

www.lean-six-sigma.pro

REGRESSION LINE

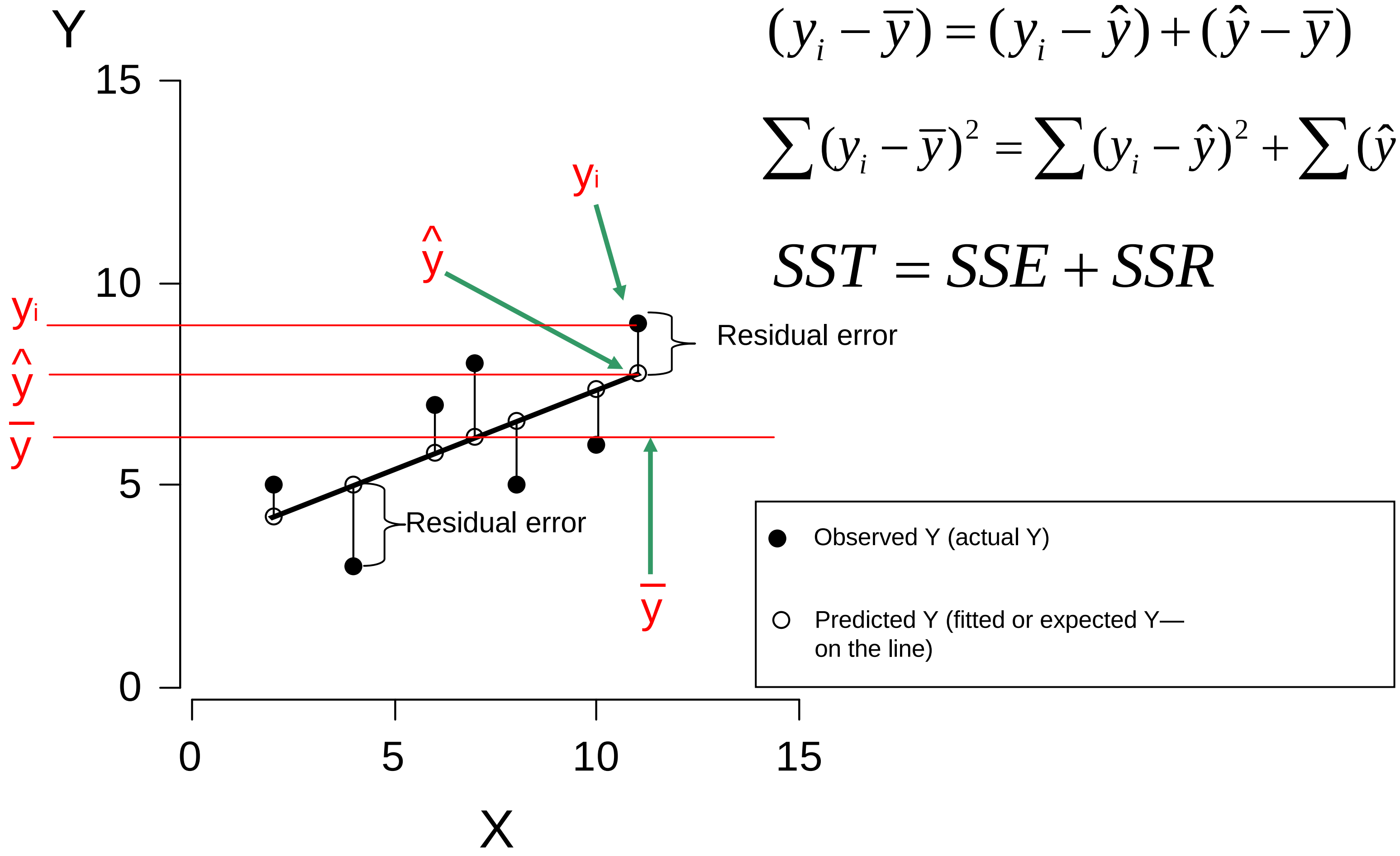


$$\hat{y} = \beta_0 + \beta_1 x$$

$$\beta_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - n\bar{x}^2}$$

We use the method of least squares to find the ***best fitting*** straight line.

REGRESSION SUMS OF THE SQUARES



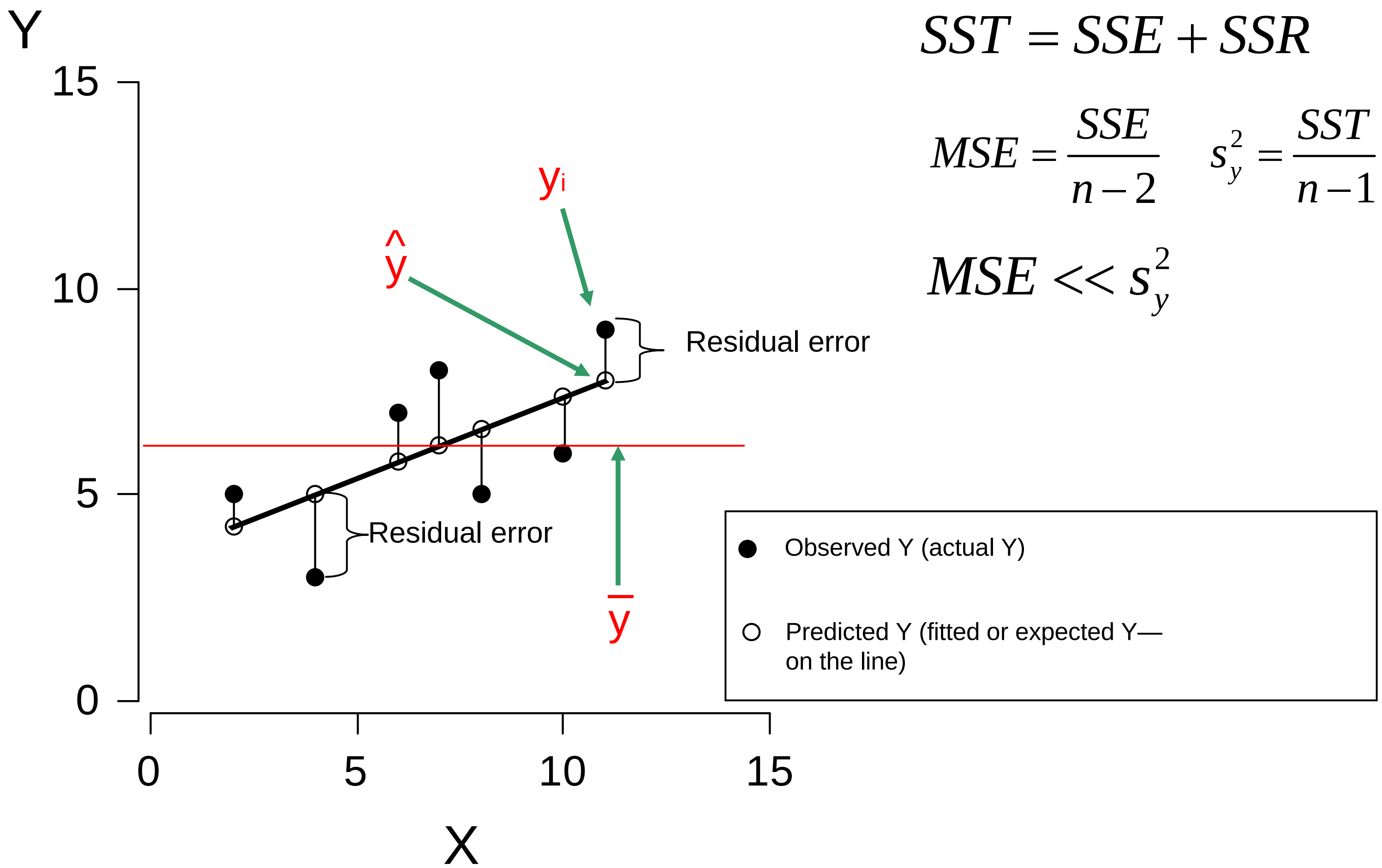
$$(y_i - \bar{y}) = (y_i - \hat{y}) + (\hat{y} - \bar{y})$$

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y})^2 + \sum (\hat{y} - \bar{y})^2$$

$$SST = SSE + SSR$$

We find the ***regression sums of squares***.

REGRESSION MEAN SQUARE ERROR AND VARIANCE OF Y



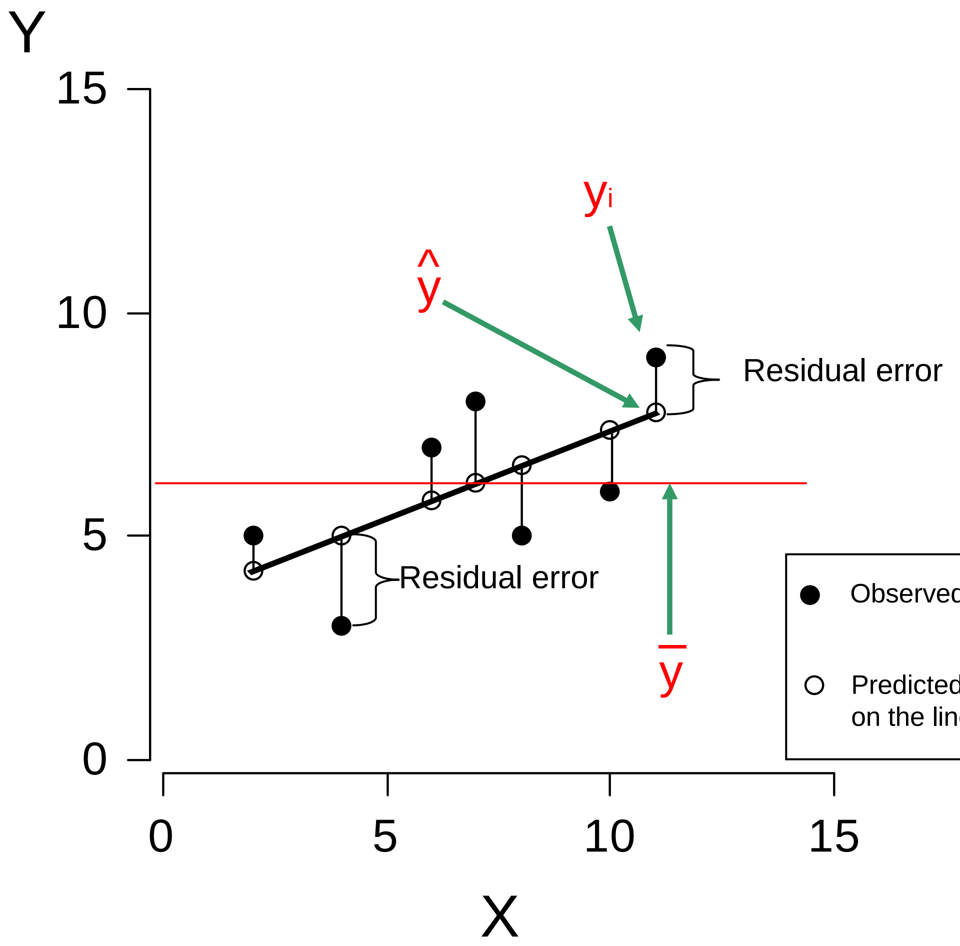
$$SST = SSE + SSR$$

$$MSE = \frac{SSE}{n-2} \qquad s_y^2 = \frac{SST}{n-1}$$

$$MSE << s_y^2$$

To find the ***best fitting*** straight line, the variance of the error (**Mean Square Error)** about the Regression Line must be **significantly less than the variance of y** about the mean.

REGRESSION AND R^2



● Observed Y (actual Y)

○ Predicted Y (fitted or expected Y—on the line)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

REGRESSION	R^2	SSE =
NONE	0	SST
PERFECT	1	0

We use R^2 to **check how much of the variability is explained by the regression line**.

$$R = \frac{\Sigma\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\Sigma\left(X_i - \bar{X}\right)^2 \Sigma\left(Y_i - \bar{Y}\right)^2}}$$

REGRESSION F STATISTIC AND HYPOTHESIS TEST

We then use the F statistic with the ANOVA table to find the p value and test the Null Hypothesis

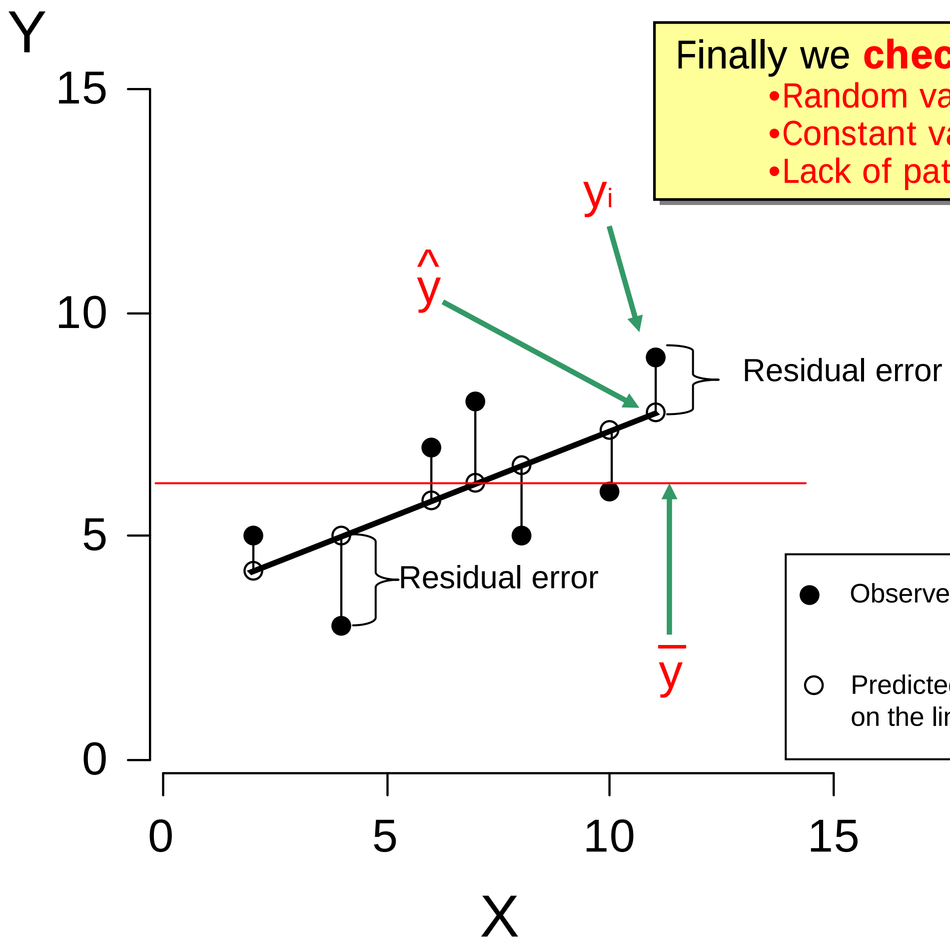
$$\hat{y} = \beta_0 + \beta_1 x$$

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

	Sum of Squares	df	Mean Square	F value
Regression	$SSR=R^2 \times SST$	1	MSR=SSR/1	MSR/MSE
Error	$SSE=(1-R^2) \times SST$	(n-2)	MSE=SSE/(n-2)	
Total	$SST=S^2_y \times (n-1)$	(n-1)		

REGRESSION



Finally we **check the residual errors** for:
- Random variation
- Constant variance
- Lack of pattern

● Observed Y (actual Y)

○ Predicted Y (fitted or expected Y—on the line)