Simple Linear Regression

Download Files

Download PDF | Download Word | Download PowerPoint

Learning Module 10: Simple Linear Regression

Variation of the Dependent Variable (Sum of Squares total)

$$
\text{Variation of }Y \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \tag{1}
$$

Where:

- Y_i: represent an observation of a company's ROA
- $\bar{Y}$: represent the mean ROA for the sample of size n
- n: sample size

View Markdown Source

```
## Variation of the Dependent Variable (Sum of Squares total)

$$
\text{Variation of }Y \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \tag{1}
$$

Where:

* $Y_i$: represent an observation of a company's ROA
```

```
* $\bar{Y}$: represent the mean ROA for the sample of size n
* $n$: sample size
```

Variation of the Independent Variable

$$
\text{Variation of } X = \sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{2}
$$

Where:

- X_i: represent an observation of our explanatory variable
- $\bar{X}$: sample mean of the independent variable
- n: number of observations

View Markdown Source

```
## Variation of the Independent Variable

$$
\text{Variation of } X = \sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{2}
$$

Where:

* $X_i$: represent an observation of our explanatory variable
* $\bar{X}$: sample mean of the independent variable
* $n$: number of observations
```

Simple Linear Regression Model

$$
Y_i = b_0 + b_1 X_i + \varepsilon_i, \dots, n. \tag{3}
$$

Where:

- Y_i: dependent variable
- X_i: independent variable
- b_0: intercept

- b_1: slope coefficient
- ε: error term

View Markdown Source

```
## Simple Linear Regression Model

$$
Y_i = b_0 + b_1 X_i + \varepsilon_i, \ldots, n. \tag{3}
$$

Where:

* $Y_i$: dependent variable
* $X_i$: independent variable
* $b_0$: intercept
* $b_1$: slope coefficient
* $\varepsilon$: error term
```

Sum of Squares Error (SSE)

$$
\begin{aligned}
\text{Sum of squares error } &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\
&= \sum_{i=1}^{n} [Y_i - (\hat{b}_0 + \hat{b}_1 X_i)]^2 \\
&= \sum_{i=1}^{n} e_i^2
\end{aligned} \tag{4}
$$

Where:

- Y_i: observed value of the dependent variable
- $\hat{Y}_i$: predicted value of the dependent variable
- $\hat{b}_0$: estimated intercept
- $\hat{b}_1$: slope coefficient
- e_i: residual for the ith observation
  - $e_i = Y_i - \hat{Y}_i$
- n: number of observations

View Markdown Source

```
## Sum of Squares Error (SSE)

$$
\text{Sum of squares error } = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2
$$

$$
= \sum_{i=1}^{n} [ Y_i - ( \hat{b}_0 + \hat{b}_1 X_i )]^2 \tag{4}
$$

$$
= \sum_{i=1}^{n} e_i^2
$$

Where:

* $Y_i$: observed value of the dependent variable
* $\hat{Y}_i$: predicted value of the dependent variable
* $\hat{b}_0$: estimated intercept
* $\hat{b}_1$: slope coefficient
* $e_i$:  residual for the $i$th observation
  * $e_i = Y_i - \hat{Y}_i$
* $n$: number of observations
```

Ordinary Least Squares Slope Estimator

$$\hat{b}_1 = \frac{\text{Covariance of } Y \text{ and } X}{\text{Variance of } X} = \frac{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{n-1}}{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

Simplifying,

$$\hat{b}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{5}$$

Where:

- $\hat{b}_1$: estimated slope coefficient

- Y_i: dependent variable observation
- X_i: independent variable observation
- $\bar{Y}$: mean of the dependent variable Y
- $\bar{X}$: mean of the independent variable X
- n: number of observations

View Markdown Source

```
## Ordinary Least Squares Slope Estimator

$$
\hat{b}_1 = \frac{\text{Covariance of } Y \text{ and } X}{\text{Variance of } X} =
\frac{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})}{n - 1}}{\frac{\sum_{i=1}^{n} (X_i
$$

Simplifying,

$$
\hat{b}_1 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{
$$

Where:

* $\hat{b}_1$: estimated slope coefficient
* $Y_i$: dependent variable observation
* $X_i$: independent variable observation
* $\bar{Y}$: mean of the dependent variable $Y$
* $\bar{X}$: mean of the independent variable $X$
* $n$: number of observations
```

Ordinary Least Squares Intercept Estimator

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} \tag{6}$$

Where:

- $\hat{b}_0$: estimated intercept
- $\bar{Y}$: mean of the dependent variable Y
- $\bar{X}$: mean of the independent variable X
- $\hat{b}_1$: estimated slope

View Markdown Source

```
## Ordinary Least Squares Intercept Estimator

$$
\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} \tag{6}
$$

Where:

* $\hat{b}_0$: estimated intercept
* $\bar{Y}$: mean of the dependent variable $Y$
* $\bar{X}$: mean of the independent variable $X$
* $\hat{b}_1$: estimated slope
```

Sample Correlation

$$
r = \frac{\text{Covariance of } Y \text{ and } X}{(\text{Standard deviation of } Y)(\text{Standard deviation of } X)} \tag{7}
$$

Where:

- r: sample correlation
- Covariance of Y and X: covariance between X and Y
- Standard deviation of X: standard deviation of the independent variable
- Standard deviation of Y: standard deviation of the dependent variable

View Markdown Source

```
## Sample Correlation

$$
r = \frac{\text{Covariance of } Y \text{ and } X}{(\text{Standard deviation of } Y)(\text{Sta
$$

Where:

* $r$: sample correlation
* $\text{Covariance of } Y \text{ and } X$: covariance between $X$ and $Y$
* $\text{Standard deviation of } X$: standard deviation of the independent variable
* $\text{Standard deviation of } Y$: standard deviation of the dependent variable
```

Homoskedasticity Assumption

$$
E(\varepsilon_i^2) = \sigma_\varepsilon^2,\quad i = 1, \ldots, n \tag{8}
$$

Where:

* ε_i:
* σ_ε^2:
* E:

View Markdown Source

```
## Homoskedasticity Assumption

$$
E(\varepsilon_i^2) = \sigma_\varepsilon^2,\quad i = 1, \ldots, n \tag{8}
$$

Where:

* $\varepsilon_i$:
* $\sigma_\varepsilon^2$:
* $E$:
```

Sum of Squares Regression (SSR)

$$
\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 \tag{9}
$$

Where:

* $\hat{Y}_i$: predicted value of the dependent variable
* $\bar{Y}$: mean of the dependent variable
* n: number of observations

View Markdown Source

```
## Sum of Squares Regression (SSR)

$$
\sum_{i=1}^{n} \left( \hat{Y}_i - \bar{Y} \right)^2 \tag{9}
$$

Where:

* $\hat{Y}_i$: predicted value of the dependent variable
* $\bar{Y}$: mean of the dependent variable
* $n$: number of observations
```

Coefficient of Determination (R^2)

$$
\text{Coefficient of determination} = \dfrac{\text{Sum of squares regression}}{\text{Sum of squares total}}
$$

$$
\text{Coefficient of determination} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \tag{10}
$$

Where:

- coefficient of determination: is the percentage of the variation of the dependent variable that is explained by the independent variable
- $\hat{Y}_i$: predicted value of the dependent variable
- Y_i: observed value of the dependent variable
- $\bar{Y}$: mean of the dependent variable
- n: number of observations

View Markdown Source

```
## Coefficient of Determination $(R^2)$

$$
\text{Coefficient of determination} = \dfrac{\text{Sum of squares regression}}{\text{Sum of s
$$

$$
\text{Coefficient of determination} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=
```

```
$$

Where:

* coefficient of determination: is the percentage of the variation of the dependent variable
* $\hat{Y}_i$: predicted value of the dependent variable
* $Y_i$: observed value of the dependent variable
* $\bar{Y}$: mean of the dependent variable
* $n$: number of observations
```

Relationship Between r^2 and R^2

$$
r^2 = \frac{\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y} \right)^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2} = R^2 \tag{11}
$$

Where:

- r: sample correlation coefficient
- R^2: coefficient of determination

View Markdown Source

```
## Relationship Between $r^2$ and $R^2$

$$
r^2 = \frac{\sum_{i=1}^{n} \left( \hat{Y}_i - \bar{Y} \right)^2}{\sum_{i=1}^{n} \left( Y_i -
$$

Where:

* $r$: sample correlation coefficient
* $R^2$: coefficient of determination
```

Mean Square Regression (MSR) with k Parameters

$$\text{MSR} = \frac{\text{Sum of squares regression}}{k} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{1} \tag{12}$$

Where:

- MSR: mean square regression
- k: number of independent variables

View Markdown Source

```
## Mean Square Regression (MSR) with k Parameters

$$
\text{MSR} = \frac{\text{Sum of squares regression}}{k} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \
$$

Where:

* $\text{MSR}$: mean square regression
* $k$: number of independent variables
```

Mean Square Regression (MSR)

$$\text{MSR} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \tag{13}$$

Where:

- MSR: mean square regression
- $\hat{Y}_i$: predicted value of the dependent variable
- $\bar{Y}$: mean of the dependent variable
- n: number of observations

View Markdown Source

```
## Mean Square Regression (MSR)

$$
\text{MSR} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \tag{13}
$$

Where:

* $\text{MSR}$: mean square regression
* $\hat{Y}_i$: predicted value of the dependent variable
* $\bar{Y}$: mean of the dependent variable
* $n$: number of observations
```

Mean Square Error (MSE)

$$
\text{MSE} = \frac{\text{Sum of squares error}}{n - k - 1}
$$

$$
\text{MSE} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n - 2} \tag{14}
$$

Where:

- MSE: mean square error
- Y_i: observed value of the dependent variable
- $\hat{Y}_i$: predicted value of the dependent variable
- n: number of observations

View Markdown Source

```
## Mean Square Error (MSE)

$$
\text{MSE} = \frac{\text{Sum of squares error}}{n - k - 1}
$$



$$
\text{MSE} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n - 2} \tag{14}
```

```
$$

Where:

* $\text{MSE}$: mean square error
* $Y_i$: observed value of the dependent variable
* $\hat{Y}_i$: predicted value of the dependent variable
* $n$: number of observations
```

F-distributed test statistic (MSR/MSE)

$$F = \frac{\frac{\text{Sum of squares regression}}{k}}{\frac{\text{Sum of squares error}}{n-k-1}} = \frac{\text{MSR}}{\text{MSE}}$$

$$F = \frac{\frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{1}}{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}} \tag{15}$$

Where:

- F: F-statistic for testing overall regression significance
- $\hat{Y}_i$: predicted value of the dependent variable
- Y_i: observed value of the dependent variable
- $\bar{Y}$: mean of the dependent variable
- $n-2$: degrees of freedom
- n: number of observations

View Markdown Source

```
## F-distributed test statistic (MSR/MSE)

$$
F = \frac{\frac{\text{Sum of squares regression}}{k}}{\frac{\text{Sum of squares error}}{n -
$$

$$
F = \frac{ \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{1}}{ \frac{\sum_{i=1}^{n} (Y_i - \ha
$$
```

```
Where:

* $F$: F-statistic for testing overall regression significance
* $\hat{Y}_i$: predicted value of the dependent variable
* $Y_i$: observed value of the dependent variable
* $\bar{Y}$: mean of the dependent variable
* $n - 2$: degrees of freedom
* $n$: number of observations
```

---

t-Test Statistic for Slope Coefficient

$$
t = \frac{\hat{b}_1 - B_1}{s_{\hat{b}_1}} \tag{16}
$$

Where:

- t: t-statistic for hypothesis test of the slope
- $\hat{b}_1$: estimated slope coefficient
- B_1: hypothesized population slope
- $s_{\hat{b}_1}$: standard error of the slope coefficient

View Markdown Source

```
## t-Test Statistic for Slope Coefficient

$$
t = \frac{\hat{b}_1 - B_1}{s_{\hat{b}_1}} \tag{16}
$$

Where:

* $t$: t-statistic for hypothesis test of the slope
* $\hat{b}_1$: estimated slope coefficient
* $B_1$: hypothesized population slope
* $s_{\hat{b}_1}$: standard error of the slope coefficient
```

Standard Error of the Slope Coefficient

$$
s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \tag{17}
$$

Where:

- $s_{\hat{b}_1}$: standard error of the slope estimate
- s_e: standard error of the estimate
- X_i:
- $\bar{X}$:
- n: number of observations

View Markdown Source

```
## Standard Error of the Slope Coefficient

$$
s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \tag{17}
$$

Where:

* $s_{\hat{b}_1}$: standard error of the slope estimate
* $s_e$: standard error of the estimate
* $X_i$:
* $\bar{X}$:
* $n$: number of observations
```

---

t-Test Statistic for Correlation

$$
t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{18}
$$

Where:

- t: t-statistic for testing correlation significance
- r: sample correlation coefficient
- n: number of observations

View Markdown Source

```
## t-Test Statistic for Correlation

$$
t = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}} \tag{18}
$$

Where:

* $t$: t-statistic for testing correlation significance
* $r$: sample correlation coefficient
* $n$: number of observations
```

Standard error of the intercept

$$
s_{\hat{b}_0} = S_e \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \tag{19}
$$

View Markdown Source

```
## Standard error of the intercept

$$
s_{\hat{b}_0} = S_e \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \
$$
```

Intercept

$$
t_{\text{intercept}} = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}} = \frac{\hat{b}_0 - B_0}{S\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}}} \tag{20}
$$

Where:

- B_0: hypothesized value

- $\hat{b}_0$: estimated intercept
- $s_{\hat{b}_0}$ standard error of the intercept

View Markdown Source

```
## Intercept

$$
t_{\text{intercept}} = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}} = \frac{\hat{b}_0 - B_0}{S \sqr
$$

Where:

* $B_0$: hypothesized value
* $\hat{b}_0$: estimated intercept
* $s_{\hat{b}_0}$ standard error of the intercept
```

Hypothesis Tests of Slope When the Independent Variable Is an Indicator Variable

$$\text{RET}_i = b_0 + b_1 \text{EARN}_i + \varepsilon_i \tag{21}$$

Where:

- RET: monthly returns
- EARN: Indicator variable

View Markdown Source

```
## Hypothesis Tests of Slope When the Independent Variable Is an Indicator Variable

$$
\text{RET}_i = b_0 + b_1 \text{EARN}_i + \varepsilon_i \tag{21}
$$

Where:

* RET: monthly returns
* EARN: Indicator variable
```

Standard Error of the Estimate

$$\text{Standard error of the estimate } (s_e) = \sqrt{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}} \tag{22}$$

Where:

- s_e: standard error of the estimate
  - The s_e is a measure of the distance between the observed values of the dependent variable and those predicted from the estimated regression—the smaller the se, the better the fit of the model

View Markdown Source

```
## Standard Error of the Estimate

$$
\text{Standard error of the estimate } (s_e) = \sqrt{\text{MSE}} = \sqrt{ \frac{\sum_{i=1}^{r
$$

Where:

* $s_e$: standard error of the estimate
  * The $s_e$ is a measure of the distance between the observed values of the dependent varia
the better the fit of the model
```

Forecasted Value of the Dependent Variable

$$\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f \tag{23}$$

Where:

- $\hat{Y}_f$: forecasted value of the dependent variable
- X_f: forecasted independent variable

View Markdown Source

```
## Forecasted Value of the Dependent Variable

$$
\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f \tag{23}
$$

Where:

* $\hat{Y}_f$: forecasted value of the dependent variable
* $X_f$: forecasted independent variable
```

Standard Error of the Forecast

estimated variance of the prediction error

$$s_f^2 = s_e^2 \left[1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{(n-1)s_X^2}\right] = s_e^2 \left[1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

Where:

s_f^2: estimated variance of the prediction error

Standard Error of the Forecast

$$s_f = s_e \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \tag{24}$$

Where:

- s_f: standard error of the forecast
- s_e: standard error of the estimate
- X_f: forecast value of the independent variable

View Markdown Source

```
## Standard Error of the Forecast

**estimated variance of the prediction error**

$$
s_f^2 = s_e^2 \left[ 1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{(n - 1)s_X^2} \right]
= s_e^2 \left[ 1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \
$$

Where:

$s_f^2$: estimated variance of the prediction error

**Standard Error of the Forecast**

$$
s_f = s_e \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}
$$

Where:

* $s_f$: standard error of the forecast
* $s_e$: standard error of the estimate
* $X_f$: forecast value of the independent variable
```

Log-Lin Model

$$\ln Y_i = b_0 + b_1 X_i \tag{25}$$

Where:

- Y_i: dependent variable
  - is in logarithmic form
- X_i: independent variable
  - not in logarithmic form
- b_0: intercept
- b_1: slope coefficient

View Markdown Source

```
## Log-Lin Model

$$
\ln Y_i = b_0 + b_1 X_i \tag{25}
$$

Where:

* $Y_i$: dependent variable
  * is in logarithmic form
* $X_i$: independent variable
  * not in logarithmic form
* $b_0$: intercept
* $b_1$: slope coefficient
```

Lin-Log Model

$$
Y_i = b_0 + b_1 \ln X_i \tag{26}
$$

Where:

- Y_i: dependent variable
  - not in logarithmic form
- X_i: independent variable
  - is in logarithmic form
- b_0: intercept
- b_1: slope coefficient

View Markdown Source

```
## Lin-Log Model

$$
Y_i = b_0 + b_1 \ln X_i \tag{26}
$$
```

```
Where:

* $Y_i$: dependent variable
  * not in logarithmic form
* $X_i$: independent variable
  * is in logarithmic form
* $b_0$: intercept
* $b_1$: slope coefficient
```

Log-Log Model

$$
\ln Y_i = b_0 + b_1 \ln X_i \tag{27}
$$

Where:

- Y_i: dependent variable
  - logarithmic form
- X_i: independent variable
  - logarithmic form
- b_0: intercept
- b_1: slope coefficient
- This model is useful in calculating elasticities because the slope coefficient is the relative change in the dependent variable for a relative change in the independent variable.

View Markdown Source

```
## Log-Log Model

$$
\ln Y_i = b_0 + b_1 \ln X_i \tag{27}
$$

Where:

* $Y_i$: dependent variable
  * logarithmic form
* $X_i$: independent variable
```

```
  * logarithmic form
* $b_0$: intercept
* $b_1$: slope coefficient
* This model is useful in calculating elasticities because the slope
  coefficient is the relative change in the dependent variable for a
  relative change in the independent variable.
```
