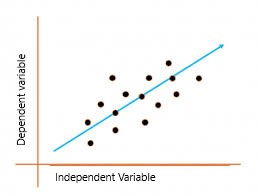
1. Using a graph to illustrate slope and intercept, define basic linear regression.

Linear regression predicts the relationship between two variables by assuming they have a straight-line connection. It finds the best line that minimizes the differences between predicted and actual values. Used in fields like economics and finance, it helps analyze and forecast data trends. Linear regression can also involve several variables (multiple linear regression) or be adapted for yes/no questions (logistic regression).

Simple Linear Regression

In a simple linear regression, there is one independent variable and one dependent variable. The model estimates the slope and intercept of the line of best fit, which represents the relationship between the variables. The slope represents the change in the dependent variable for each unit change in the independent variable, while the intercept represents the predicted value of the dependent variable when the independent variable is zero.

Linear regression is a quiet and the simplest statistical regression technique used for predictive analysis in machine learning. Linear regression shows the linear relationship between the independent(predictor) variable i.e. X-axis and the dependent (output) variable i.e. Y-axis, called linear regression. If there is a single input variable X (independent variable), such linear regression is ***simple linear regression***.



The graph above presents the linear relationship between the output(y) and predictor(X) variables. The blue line is referred to as the best-fit straight line. Based on the given data points, we attempt to plot a line that fits the points the best.

Simple Regression Calculation

To calculate best-fit line linear regression uses a traditional slope-intercept form which is given below,

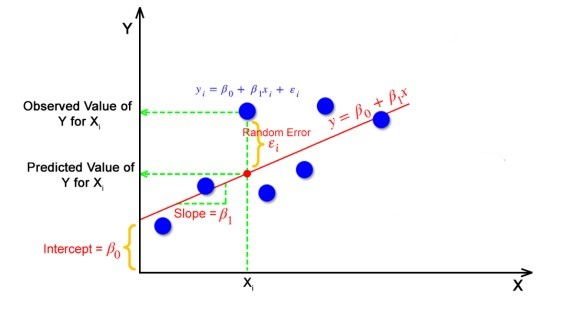
**Y i =  β 0 + β 1 X i**

where Y i  = Dependent variable,  **β 0** = constant/Intercept, **β 1** = Slope/Intercept, **X i** = Independent variable.

This algorithm explains the linear relationship between the dependent(output) variable y and the independent(predictor) variable X using a straight line  Y= B 0 + B 1 X.

But how does the linear regression find out which is the best-fit line?

The goal of the linear regression algorithm is to get the **best values for B 0 and B 1** to find the best-fit line. The best-fit line is a line that has the least error which means the error between predicted values and actual values should be minimum.



But how the linear regression finds out which is the best fit line?

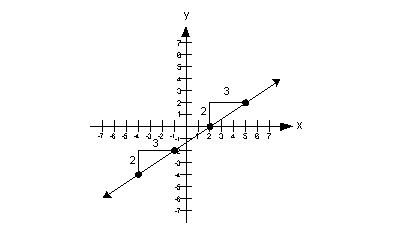
The goal of the linear regression algorithm is to get the **best values for B0 and B1** to find the best fit line. The best fit line is a line that has the least error which means the error between predicted values and actual values should be minimum.

1. In a graph, explain the terms rise, run, and slope.

The slope of a line **measures the steepness of the line.**

Most of you are probably familiar with associating slope with "rise over run". 

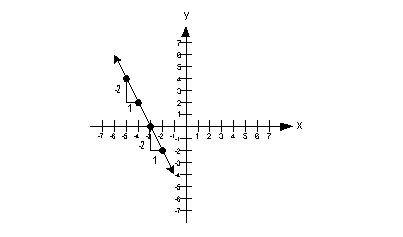
**Rise** means how many units you move up or down from point to point.  On the graph that would be a **change in the *y* values.**

**Run**means how far left or right you move from point to point.  On the graph, that would mean a **change of *x* values**.  
**Here are some visuals to help you with this definition:**  
**Positive slope:**  


positive slope

**Note that when a line has a positive slope it goes up left to right.**

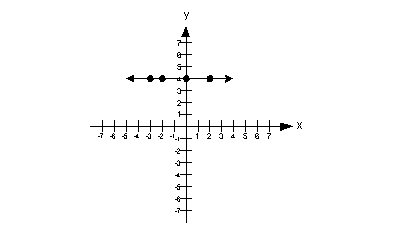
**Negative slope:**



negative slope

**Note that when a line has a negative slope it goes down left to right.**

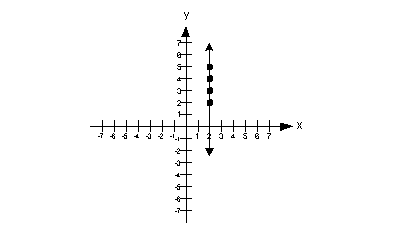
**Zero slope:**



*slope* = 0

**Note that when a line is horizontal the slope is 0.**

**Undefined slope:**



*slope* = undefined

**Note that when the line is vertical the slope is undefined.**  
   
**Here is a little review on the slope/intercept form of the equation of a line.**  If you need more of a review on slope, feel free to go to [**Tutorial 26: Equations of Lines**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut26_eqline.htm).

**Slope/Intercept Equation of a Line**

y intercept  
 If your linear equation is written in this form, ***m* represents the slope and *b* represents the *y*-intercept.**

This form can be handy if you need to find the slope of a line given the equation.  
   
**Graphing a Line Using the *y*-intercept and Slope**  
**Step 1:** **Put the equation in slope/intercept form (**y intercept**) and identify the slope and *y*-intercept.**

When it is in this form it makes it easier to identify the slope and the *y*-intercept of the line.

In this form, the slope is *m*, which is the coefficient in front of *x* and the *y*-intercept is the constant *b*.

**Step 2:** **Plot the *y*-intercept on a 2-dimensional graph.**

Recall that the *y*-intercept is where it crosses the *y*-axis.  So *x*'s value will be 0.

If you need more review on intercepts, feel free to go to [**Tutorial 26: Equations of Lines**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut26_eqline.htm).

**Step 3:** **Use the slope to find a second point on the line.**

You will start with the*y*-intercept and rise up or down and then run left or right depending on the sign of the slope.

Use the concept of slope being rise over run (rise/run) to determine how to use the slope to find your second point.

**What do you think is the positive direction for rise, up or down?**   If you said up, pat yourself on the back.  That means down is the negative direction for rise.  A way to remember this is to think about the numbers are on the *y*-axis.  Going up above the origin are the positive values, and going down below the origin are the negative values.

**What do you think is the positive direction for run, right or left?**  If you said right, pat yourself on the back.  That means left is the negative direction for run.  A way to remember this is to think about the numbers are on the*x*-axis.  Going to the right of  the origin are the positive values, and going to the left of the origin are the negative values. 

**If the slope is positive**, then the rise and the run need to either be BOTH positive or BOTH negative.  In other words, you will be **going up and to the right OR down and to the left**.  The reason both negative directions work is our slope is rise over run and if you have a negative over a negative, it simplifies to be a positive.

**If the slope is negative**, then the rise and the run have to be opposites of each other, one has to be positive and one has to be negative.  In other words, you will be **going up and to the left OR down and to the right**.  If you make them both negative, then it would simplify to be a positive and you would have the wrong graph and you don't want to do that. 

Also keep in mind, if your slope is a non-zero integer like -5 or 10, that there is a denominator or run of your slope.  **What is the denominator of a non-zero integer?**  If you said 1, you are correct!!!

**Step 4:** **Draw a line through the points found in steps 2 and 3.**

All of the graphs in this tutorial will be straight lines.

**Example 1**:  Give the slope and *y*-intercept of the line example 1a and then graph it  
**Step 1:** [**Put the equation in slope/intercept form**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step1)**(**y intercept**) and identify the slope and *y*-intercept.**

This equation is already in slope/intercept form:

example 1b

**\*Slope/intercept form**

Lining up the form with the equation we have been given, can you see what the slope and *y-*intercept are?

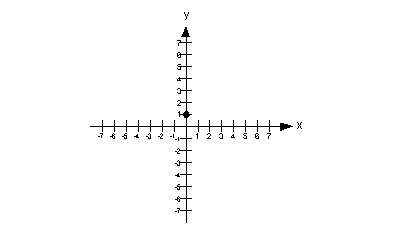
In this form, the **slope is *m***, which is the number in front of *x*.  In our problem that **would have to be 3.**

In this form, the ***y*-intercept is *b***, which is the constant.  In our problem **that would be 1**.

How did you do? 

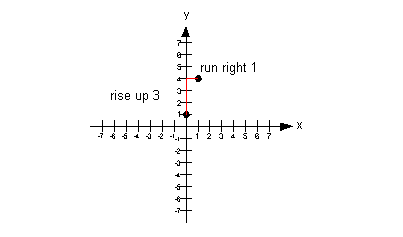
**Step 2:** [**Plot the *y*-intercept on a 2-dimensional graph.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step2)

Since the *y*-intercept is where the line crosses the *y*-axis, then *x*'s value would have to be 0.  In step 1 we found our *y*-intercept value to be 1.

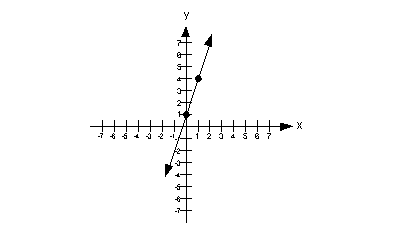
**Putting that together, the ordered pair for the *y*-intercept  would be (0, 1):**  
    
   
**Step 3:** [**Use the slope to find a second point on the line.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step3)

In step 1 we found our slope to be 3.  What would the denominator of 3 be?  If you said 1, you are right on.  So we can think of 3 as 3/1.  That makes it easier to think of it as rise over run (rise/run).

Since we have a positive slope, the rise and the run need to either be BOTH positive or BOTH negative.  So, we can either rise up 3 and run right 1 OR go down 3 and left 1.

I chose to **rise up 3 and run right 1, starting on the *y*-intercept::**  
  

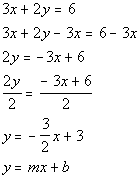
**Note that if we would have gone down 3 and left 1 from our *y*-intercept, that we would have ended up at (-1, -2) which would have lined up with the other points.**  
   
**Step 4:** [**Draw a line through the points found in steps 2 and 3.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step4)

**Example 2**:  Give the slope and *y*-intercept of the line example 2a and then graph it.

**Step 1:** [**Put the equation in slope/intercept form**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step1)**(**y intercept**) and identify the slope and *y*-intercept.**

**Solve for *y* to get it into the slope/intercept form:**



**\*Inverse of add. 3*x* is sub. 3*x***

**\*Inverse of mult. 2 is div. 2**

**\*Slope/intercept form**

 Lining up the form with the equation we got, can you see what the slope and *y-*intercept are?

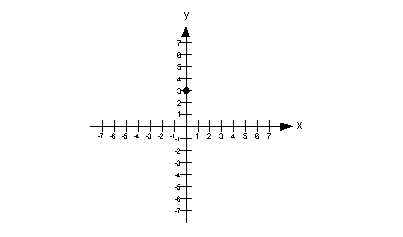
In this form, the **slope is *m***, which is the number in front of *x*.  In our problem that **would have to be -3/2.**

In this form, the ***y*-intercept is *b***, which is the constant.  In our problem **that would be 3**.

How did you do? 

**Step 2:** [**Plot the *y*-intercept on a 2-dimensional graph.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step2)

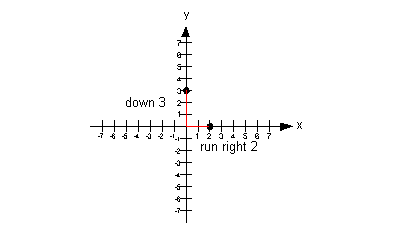
Since the *y*-intercept is where the line crosses the *y*-axis, then *x*'s value would have to be 0.  In step 1 we found our *y*-intercept value to be 3.

**Putting that together, the ordered pair for the *y*-intercept  would be (0, 3):**  
    
 

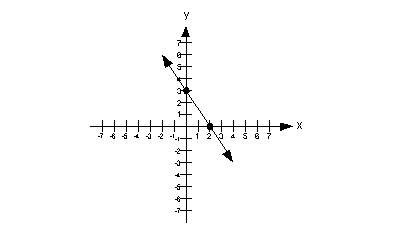
**Step 3:** [**Use the slope to find a second point on the line.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step3)

In step 1 we found our slope to be -3/2.

Since we have a negative slope the rise and the run have to be opposites of each other, one has to be positive and one has to be negative.  So, we can either go down 3 and run right 2 OR rise up 3 and run left 2.

I chose to **go down 3 and run right 2, starting on the *y*-intercept:**  
  

**Note that if we rose up 3 and ran left 2 from our *y*-intercept,  we would have ended up at (-2, 6) which would have lined up with the other points.**  
   
**Step 4:** [**Draw a line through the points found in steps 2 and 3.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step4)

**Example 3**:  Give the slope and *y*-intercept of the line example 3a and then graph it.

**Step 1:** [**Put the equation in slope/intercept form**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step1)**(**y intercept**) and identify the slope and *y*-intercept.**

This equation is already in slope/intercept form:



**\*Slope/intercept form**

Lining up the form with the equation we have been given, can you see what the slope and *y-*intercept are?

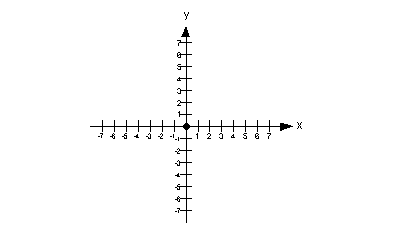
In this form, the **slope is *m***, which is the number in front of *x*.  In our problem that **would have to be 5/2.**

In this form, the ***y*-intercept is *b***, which is the constant.  In our problem **that would be 0**.

How did you do? 

**Step 2:** [**Plot the *y*-intercept on a 2-dimensional graph.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step2)

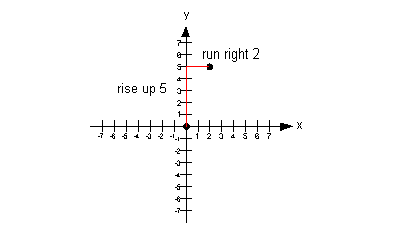
Since the *y*-intercept is where the line crosses the *y*-axis, then *x*'s value would have to be 0.  In step 1 we found our *y*-intercept value to be 0.

**Putting that together, the ordered pair for the *y*-intercept  would be (0, 0):**  
    
 

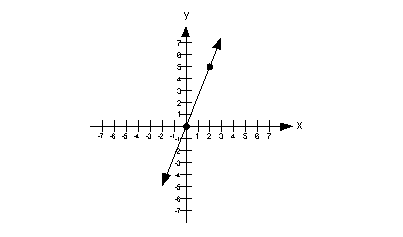
**Step 3:** [**Use the slope to find a second point on the line.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step3)

In step 1 we found our slope to be 5/2.

Since we have a positive slope the rise and the run need to either be BOTH positive or BOTH negative.  So, we can either rise up 5 and run right 2 OR go down 5 and left 2.

I chose to **rise up 5 and run right 2, starting on the *y*-intercept:**  
  

**Note that if we would have gone down 5 and left 2 from our *y*-intercept, we would have ended up at (-2, -5) which would have lined up with the other points.**  
   
**Step 4:** [**Draw a line through the points found in steps 2 and 3.**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#step4)

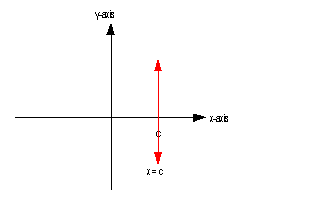
  

**Vertical Lines**

***x*= *c***

If you have an equation*x* = *c*, where *c* is a constant, and you are wanting to graph it on a two dimensional graph, this would be a vertical line with *x*-intercept of (*c*, 0).

Even though you do not see a *y*in the equation, you can still graph it on a two dimensional graph.  Remember that the graph is the set of all solutions for a given equation.  If all the points are solutions, then any ordered pair that has an *x* value of *c* would be a solution.  As long as  *x*never changes value,  it is always *c*, then you have a solution.  In that case, you will end up with a vertical line.   
   
**Below is an illustration of a vertical line *x* = *c*:**



As mentioned above, [**the slope of a vertical line is undefined**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#vertical).

Also, note that except for the vertical line *x* = 0, a vertical line does not go through the *y*-axis.  So that means **a vertical line has no *y*-intercept, unless it is *x* = 0.**

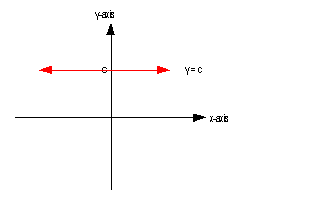
**Horizontal Lines**

***y*= *c***

If you have an equation *y* = *c*, where *c* is a constant, and you are wanting to graph it on a two dimensional graph, this would be a horizontal line with *y*- intercept of (0, *c*).

Even though you do not see an *x*in the equation, you can still graph it on a two dimensional graph.  Remember that the graph is the set of all solutions for a given equation.  If all the points are solutions, then any ordered pair that has a *y* value of *c* would be a solution.  As long as y never changes value,  it is always c, then you have a solution.  In that case, you will end up with a horizontal line.

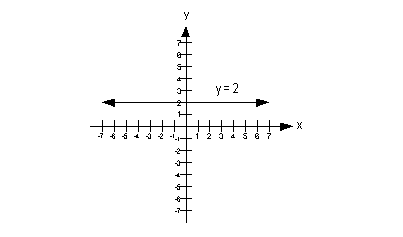
**Below is an illustration of a horizontal line *y* = *c*:**



As mentioned above, [**a horizontal line has a slope of 0**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#horizontal).

A horizontal line's *y*-intercept is whatever *y* is set equal to.  
   
**Example 4**:  Give the slope and *y*-intercept of the line example 4a and then graph it.

Note how we have an equation that has *y* set equal to a constant, where there is no *x*.  Since this [**fits the form of *y* =*c* described above**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#horizontal2), we can cut to the chase and **draw our horizontal line:**

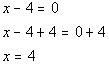
  

Since we have a horizontal line, what is our slope going to be?  If you said 0, you are so right!!!

What would the *y*-intercept be?  Give yourself a high five if you said 2.  
   
**Example 5**:  Give the slope and *y*-intercept of the line example 5a and then graph it.

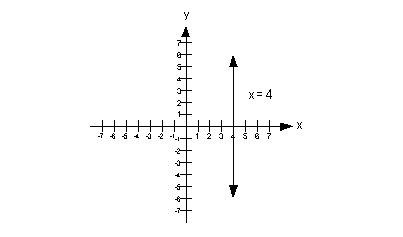
Note how we are missing a *y* and even though it is not quite in the form *x* = *c*, that we can get it in that form.

**Lets first rewrite this in the form *x* = *c*and then go from there:**



**\*Written in the form *x*= *c***

Note how we have an equation that has *x* set equal to a constant, where there is no *y*.  Since this [**fits the form of *x* =*c* described above**](https://www.wtamu.edu/academic/anns/mps/math/mathlab/col_algebra/col_alg_tut27_graphline.htm#vertical2), we can jump to the chase and **draw our vertical line:**

3. Use a graph to demonstrate slope, linear positive slope, and linear negative slope, as well as the different conditions that contribute to the slope.

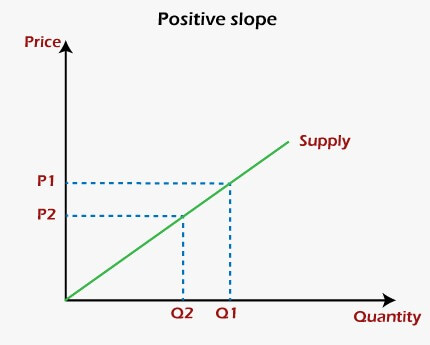
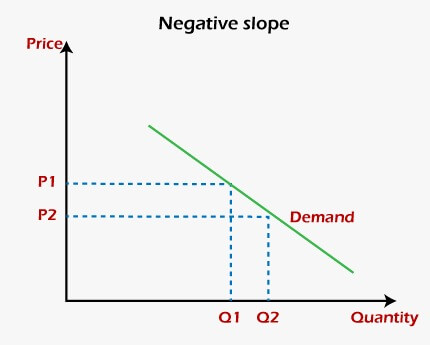
WHAT IS A SLOPE?

A number of absolute values that represent whether a line is **steeper or flatter** and the **direction** of the line on the graph are known as a **slope or gradient.** The slope of a line is fundamental concept in economics and mathematics. It is generally denoted by the letter **'m'.** The slope can be calculated by dividing the **'vertical change'** with the **'horizontal change'** between two distinct points on a line.

TYPES OF SLOPE

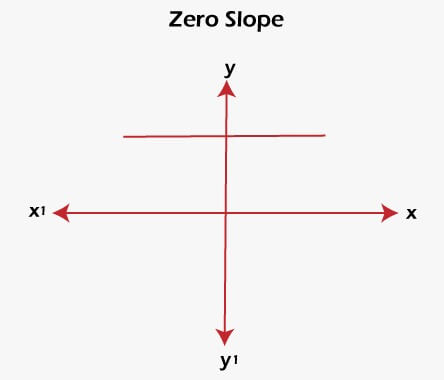
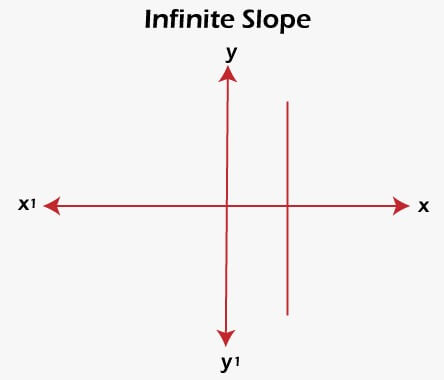
There are two main types of slopes which are given below:

ADVERTISEMENT

* **Positive Slope:** A slope in which two variables, i.e., variable at x-axis and variable at the y-axis, shows a **positive relationship** is known as **positive slope.** In simpler words, a positive slope is one in which the variable x increases with the increase in variable y and/or variable y increases with the increase in variable x. Similarly, the variable x decreases with the decrease in variable y, and/or variable y decreases with the decrease in variable x. It means both the variables are **complements** to each other. A positive slope moves in the **upward direction** or is **upward sloping.**  
  In graphical terms, a positive slope is one in which the line on the graph rises when it moves from left to right. The concept of positive slope can be clearly understood with the help of the **supply curve** of a producer or firm in economics. The two variables of the curve are price at the y-axis and quantity of goods at the x-axis. Let us assume the firm is producing the goods for **profit maximization.** Therefore, when the prices of the goods increase, the quantity supplied by the firm of those goods will also increase, while when the prices decrease, the quantity supplied by the firm will decrease. In other words, at higher prices, the firm or producer will increase the quantity supplied to earn more profit, while at lower prices, they will reduce the quantity supplied to reduce the loss. Hence, it shows the prices and quantity supplied are positively related to each other, which can be cleared from the diagram given below:  
  
* **Negative Slope:** A slope in which two variables, i.e., variable at x-axis and variable at the y-axis, shows a **negative relationship** is known as **negative slope.** In other words, a negative slope is one in which the variable x increases with the decrease in variable y and/or variable y increases with the decrease in variable x. In the same manner, the variable x decreases with the increase in variable y, and/or variable y decreases with the increase in variable x. This represents an **inverse relationship** between these two variables. A negative slope moves in the **downward direction** or is **downward sloping.**  
  Graphically, a negative slope is one in which the line on the graph falls when it moves from left to right. One of the best examples of the negative slope of the graph is the **demand curve** in economics. The two variables of the curve are price at the y-axis and quantity of goods at the x-axis. As we know, the consumers buy a large quantity of a good at a lower price than at a higher price. Therefore, the quantity demanded by the consumers of goods will decrease with an increase in the prices of those goods. On the other hand, when prices of the goods will decrease, the quantity demand will increase. Hence, it shows a **negative relationship** between the prices and quantity supplied of those goods. It can be cleared from the diagram given below:  
  

TWO OTHER TYPES OF SLOPE

Other than positive and negative slopes, there are two more types of slopes named zero slope and infinite slope. They can be understood from the given explanation:

* **Zero Slope:** A condition in which the variable at the y-axis remains the same with the change in the variable at the x-axis is known as the **slope of zero.** Graphically, a **horizontal or flat line** on the graph has a zero Slope. Hence, it is called a **constant function.** A slope of zero neither moves into the upward or downward direction. It moves only to the leftward or rightward directions.  
  The diagram given below is a graphical presentation of the zero slope:  
  
* **Infinite Slope:** A condition in which the variable at the x-axis remains the same with the change in the variable at the y-axis is known as the **infinite slope.** It is also called the **undefined slope.** As per the graphical terms, a **vertical or perpendicular line** on the graph has an infinite slope. An infinite slope neither moves to the leftward or rightward direction. It shows a movement only in the upward or downward direction.  
  An infinite slope is shown in the given diagram:  
  

CALCULATION OF SLOPE

* In a linear equation of **ax + by + c = 0,** the slope is defined as **-a/b.**
* The equation of the line can be calculated with the help of the **point-slope formula** if both the slope **m** of a line and point **(x1, y1)** are known. The formula is given below:  
  **y - y1 = m (x - x1)**
* The two-line will be **parallel** if their slopes are **equal,** while two lines will be **perpendicular** if the product of their slopes is **-1.**

ADDITIONAL INFORMATION

* The absolute value of the slope is used to find whether a curve is **steeper or flatter.**
* The positive and negative value of the slope decides the direction, i.e., **upward or downward,** of the slope.
* A curve becomes **steeper** with the increase in the absolute value of the slope.
* A curve becomes **flattered** with the decrease in the absolute value of the slope.
* These conditions are not affected by the **negative or positive slope** (not the negative or positive value).
* A **lower positive slope** implies a flatter curve that is tilted in the upward direction will be formed.
* A **higher positive slope** means a steeper curve that is bent in the upward direction will be formed.
* A **negative slope with a large absolute value** implies a steeper curve that is tilted in the downward direction will be formed.
* A **negative slope having a smaller absolute value** means a flatter curve that is bent in the downward direction will be formed.

4. Use a graph to demonstrate curve linear negative slope and curve linear positive slope.

* The absolute value of the slope is used to find whether a curve is **steeper or flatter.**
* The positive and negative value of the slope decides the direction, i.e., **upward or downward,** of the slope.
* A curve becomes **steeper** with the increase in the absolute value of the slope.
* A curve becomes **flattered** with the decrease in the absolute value of the slope.
* These conditions are not affected by the **negative or positive slope** (not the negative or positive value).
* A **lower positive slope** implies a flatter curve that is tilted in the upward direction will be formed.
* A **higher positive slope** means a steeper curve that is bent in the upward direction will be formed.
* A **negative slope with a large absolute value** implies a steeper curve that is tilted in the downward direction will be formed.
* A **negative slope having a smaller absolute value** means a flatter curve that is bent in the downward direction will be formed.

5. Use a graph to show the maximum and low points of curves.

Let’s say, the maxima/minima point(s) of a function y=f(x) needs to be identified. Then the required steps are as outlined below.

# Evaluate the derivative of the function (y’ or f’(x)) and equate it to zero. This gives the values of x, say x[i] in general, where the given function has stationary point(s), viz the slope of tangent is zero, means the tangent at the point(s) is/are parallel to the horizontal (X-)axis.

# Obtain the second derivative (y’’ or f’’(x)) and substitute the values of x, obtained above, by turns. If f’’(x[i])<0, the function has point(s) of maxima while for f’’(x[i])>0, the function has point(s) of minima.

6. Use the formulas for a and b to explain ordinary least squares.

In [statistics](https://en.wikipedia.org/wiki/Statistics), **ordinary least squares** (**OLS**) is a type of [linear least squares](https://en.wikipedia.org/wiki/Linear_least_squares) method for choosing the unknown [parameters](https://en.wikipedia.org/wiki/Statistical_parameter) in a [linear regression](https://en.wikipedia.org/wiki/Linear_regression) model (with fixed level-one[[*clarification needed*](https://en.wikipedia.org/wiki/Wikipedia:Please_clarify)] effects of a [linear function](https://en.wikipedia.org/wiki/Linear_function) of a set of [explanatory variables](https://en.wikipedia.org/wiki/Explanatory_variable)) by the principle of [least squares](https://en.wikipedia.org/wiki/Least_squares): minimizing the sum of the squares of the differences between the observed [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable) (values of the variable being observed) in the input [dataset](https://en.wikipedia.org/wiki/Dataset) and the output of the (linear) function of the [independent variable](https://en.wikipedia.org/wiki/Independent_variable). Some sources consider OLS to be linear regression.[[1]](https://en.wikipedia.org/wiki/Ordinary_least_squares#cite_note-1)

Geometrically, this is seen as the sum of the squared distances, parallel to the axis of the dependent variable, between each data point in the set and the corresponding point on the regression surface—the smaller the differences, the better the model fits the data. The resulting [estimator](https://en.wikipedia.org/wiki/Statistical_estimation) can be expressed by a simple formula, especially in the case of a [simple linear regression](https://en.wikipedia.org/wiki/Simple_linear_regression), in which there is a single [regressor](https://en.wikipedia.org/wiki/Regressor" \o "Regressor) on the right side of the regression equation.

The OLS estimator is [consistent](https://en.wikipedia.org/wiki/Consistent_estimator) for the level-one fixed effects when the regressors are [exogenous](https://en.wikipedia.org/wiki/Exogenous) and forms perfect [colinearity](https://en.wikipedia.org/wiki/Collinearity" \o "Collinearity) (rank condition), consistent for the variance estimate of the residuals when regressors have finite fourth moments[[2]](https://en.wikipedia.org/wiki/Ordinary_least_squares#cite_note-2) and—by the [Gauss–Markov theorem](https://en.wikipedia.org/wiki/Gauss%E2%80%93Markov_theorem)—[optimal in the class of linear unbiased estimators](https://en.wikipedia.org/wiki/Best_linear_unbiased_estimator) when the [errors](https://en.wikipedia.org/wiki/Statistical_error) are [homoscedastic](https://en.wikipedia.org/wiki/Homoscedastic" \o "Homoscedastic) and [serially uncorrelated](https://en.wikipedia.org/wiki/Autocorrelation). Under these conditions, the method of OLS provides [minimum-variance mean-unbiased](https://en.wikipedia.org/wiki/UMVU) estimation when the errors have finite [variances](https://en.wikipedia.org/wiki/Variance). Under the additional assumption that the errors are [normally distributed](https://en.wikipedia.org/wiki/Normal_distribution) with zero mean, OLS is the [maximum likelihood estimator](https://en.wikipedia.org/wiki/Maximum_likelihood_estimator) that outperforms any non-linear unbiased estimator

7. Provide a step-by-step explanation of the OLS algorithm.

Applying the OLS method is a systematic process and involves several steps: **Step 1:** Gather data for the variables you are interested in. **Step 2:** Plot these data points on a scatter plot with the dependent variable on the y-axis and the independent variable on the x-axis. **Step 3:** Use the OLS formula to calculate the slope (𝑏) and y-intercept (𝑎) of the regression line. **Step 4:** Draw the regression line on the scatter plot using the slope and y-intercept. **Step 5:** Use this line to predict the dependent variable's value for different independent variable values. It's important to remember that even though OLS regression can provide insight into relationships between variables, correlation does not equate to causation. Other factors may influence the observed relationships. By becoming well-versed in the OLS method, you are equipping yourself with a powerful tool for making [**data-driven decisions**](https://www.studysmarter.co.uk/explanations/business-studies/managerial-economics/data-driven-decisions/) in business studies. It offers a way to quantify risk, forecast future results, and understand the impact of various factors on a desired outcome.

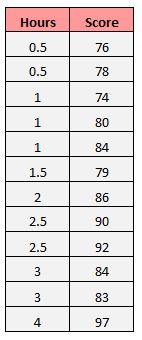
8. What is the regression's standard error? To represent the same, make a graph.

When we fit a [regression model](https://www.statology.org/introduction-to-simple-linear-regression/) to a dataset, we’re often interested in how well the regression model “fits” the dataset. Two metrics commonly used to measure goodness-of-fit include [R-squared](https://www.statology.org/what-is-a-good-r-squared-value/) (R2) and **the** **standard error of the regression**, often denoted *S*.

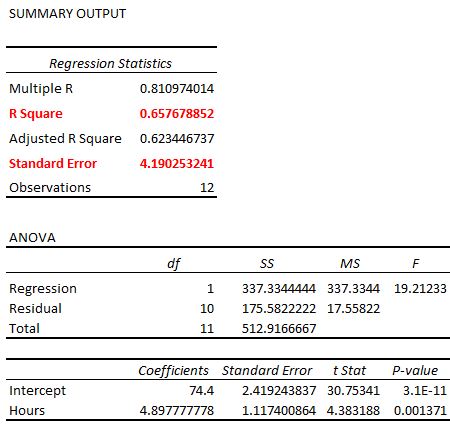
This tutorial explains how to interpret the standard error of the regression (S) as well as why it may provide more useful information than R2.

**Standard Error vs. R-Squared in Regression**

Suppose we have a simple dataset that shows how many hours 12 students studied per day for a month leading up to an important exam along with their exam score:



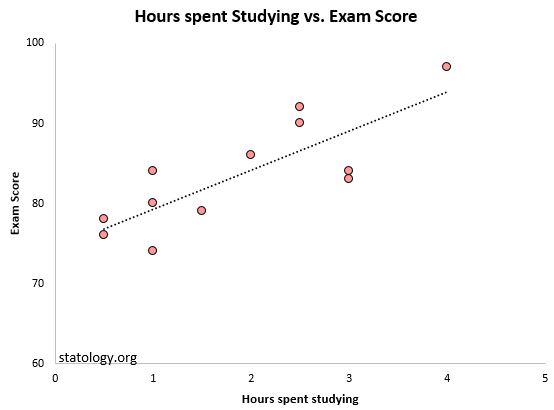
If we fit a simple linear regression model to this dataset in Excel, we receive the following output:



**R-squared** is the proportion of the variance in the response variable that can be explained by the predictor variable. In this case, **65.76%** of the variance in the exam scores can be explained by the number of hours spent studying.

**The standard error of the regression** is the average distance that the observed values fall from the regression line. In this case, the observed values fall an average of 4.89 units from the regression line.

If we plot the actual data points along with the regression line, we can see this more clearly:

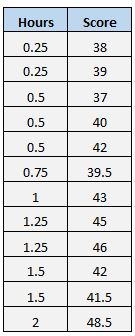


Notice that some observations fall very close to the regression line, while others are not quite as close. But on average, the observed values fall**4.19 units** from the regression line.

The standard error of the regression is particularly useful because it can be used to assess the precision of predictions. Roughly 95% of the observation should fall within +/- two standard error of the regression, which is a quick approximation of a 95% prediction interval.

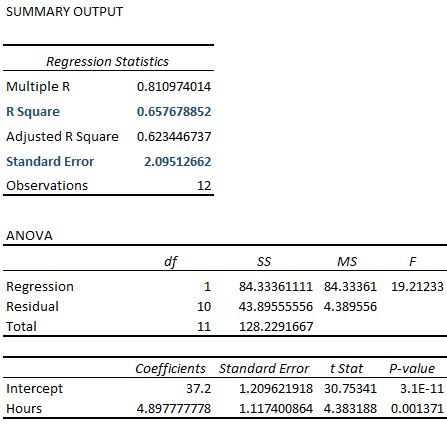
If we’re interested in making predictions using the regression model, the standard error of the regression can be a more useful metric to know than R-squared because it gives us an idea of how precise our predictions will be in terms of units.

To illustrate why the standard error of the regression can be a more useful metric in assessing the “fit” of a model, consider another example dataset that shows how many hours 12 students studied per day for a month leading up to an important exam along with their exam score:



Notice that this is the exact same dataset as before, **except all of the values are cut in half**. Thus, the students in this dataset studied for exactly half as long as the students in the previous dataset and received exactly half the exam score.

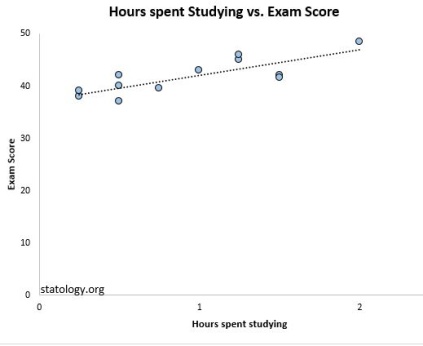
If we fit a simple linear regression model to this dataset in Excel, we receive the following output:



Notice that the R-squared of **65.76%** is the exact same as the previous example.

However, the standard error of the regression is **2.095**, which is exactly half as large as the standard error of the regression in the previous example.

If we plot the actual data points along with the regression line, we can see this more clearly:



Notice how the observations are packed much more closely around the regression line.  On average, the observed values fall**2.095 units** from the regression line.

So, even though both regression models have an R-squared of **65.76%**, we know that the second model would provide more precise predictions because it has a lower standard error of the regression.

**The Advantages of Using the Standard Error**

The standard error of the regression (S) is often more useful to know than the R-squared of the model because it provides us with actual units. If we’re interested in using a regression model to produce predictions, S can tell us very easily if a model is precise enough to use for prediction.

For example, suppose we want to produce a 95% prediction interval in which we can predict exam scores within 6 points of the actual score.

Our first model has an R-squared of 65.76%, but this doesn’t tell us anything about how precise our prediction interval will be. Luckily we also know that the first model has an S of 4.19. This means a 95% prediction interval would be roughly 2\*4.19 = +/- 8.38 units wide, which is too wide for our prediction interval.

Our second model also has an R-squared of 65.76%, but again this doesn’t tell us anything about how precise our prediction interval will be. However, we know that the second model has an S of 2.095. This means a 95% prediction interval would be roughly 2\*2.095= +/- 4.19 units wide, which is less than 6 and thus sufficiently precise to use for producing prediction intervals.

9. Provide an example of multiple linear regression.

**Multiple linear regression** is used to estimate the relationship between **two or more independent variables**and**one dependent variable**. You can use multiple linear regression when you want to know:

1. How strong the relationship is between two or more [independent variables](https://www.scribbr.com/methodology/independent-and-dependent-variables/#independent) and one dependent variable (e.g. how rainfall, temperature, and amount of fertilizer added affect crop growth).
2. The value of the [dependent variable](https://www.scribbr.com/methodology/independent-and-dependent-variables/#dependent) at a certain value of the independent variables (e.g. the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition).

10. Describe the regression analysis assumptions and the BLUE principle.

The Gauss-Markov theorem states that if your linear [regression](https://statisticsbyjim.com/glossary/regression-analysis/) model satisfies the first six classical assumptions, then [ordinary least squares](https://statisticsbyjim.com/glossary/ordinary-least-squares/) (OLS) regression produces unbiased estimates that have the smallest variance of all possible linear [estimators](https://statisticsbyjim.com/glossary/estimator/).

The proof for this theorem goes way beyond the scope of this blog post. However, the critical point is that when you satisfy the classical assumptions, you can be confident that you are obtaining the best possible [coefficient](https://statisticsbyjim.com/glossary/regression-coefficient/) estimates. The Gauss-Markov theorem does not state that these are just the best possible estimates for the OLS procedure, but the best possible estimates for *any*  linear model estimator. Think about that!

In my post about the [classical assumptions of OLS linear regression](https://statisticsbyjim.com/regression/ols-linear-regression-assumptions/), I explain those assumptions and how to verify them. In this post, I take a closer look at the nature of OLS estimates. What does the Gauss-Markov theorem mean exactly when it states that OLS estimates are the best estimates when the assumptions hold true?

The Gauss-Markov Theorem: OLS is BLUE!

The Gauss-Markov theorem famously states that OLS is BLUE. BLUE is an acronym for the following:

Best Linear Unbiased Estimator

In this context, the definition of “best” refers to the minimum variance or the narrowest sampling distribution. More specifically, when your model satisfies the assumptions, OLS coefficient estimates follow the tightest possible sampling distribution of unbiased estimates compared to other linear estimation methods.

Let’s dig deeper into everything that is packed into that sentence!

What Does OLS Estimate?

Regression analysis is like any other [inferential methodology](https://statisticsbyjim.com/basics/descriptive-inferential-statistics/). Our goal is to draw a [random sample](https://statisticsbyjim.com/glossary/sample/) from a [population](https://statisticsbyjim.com/glossary/population/) and use it to estimate the properties of that population. In regression analysis, the [coefficients in the equation](https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/) are estimates of the actual population [parameters](https://statisticsbyjim.com/glossary/parameter/).

The notation for the model of a population is the following:

Regression model notation for a population.

The betas (β) represent the population parameter for each term in the model. Epsilon (ε) represents the random error that the model doesn’t explain. Unfortunately, we’ll never know these population values because it is generally impossible to measure the entire population. Instead, we’ll obtain estimates of them using our random sample.

The notation for an estimated model from a random sample is the following:

Regression model notation for sample data.

The hats over the betas indicate that these are parameter estimates while e represents the [residuals](https://statisticsbyjim.com/regression/check-residual-plots-regression-analysis/), which are estimates of the random error.

Typically, [statisticians](https://statisticsbyjim.com/glossary/statistics/) consider estimates to be useful when they are unbiased (correct on average) and precise (minimum variance). To apply these concepts to parameter estimates and the Gauss-Markov theorem, we’ll need to understand the sampling distribution of the parameter estimates.

Sampling Distributions of the Parameter Estimates

Imagine that we repeat the same study many times. We collect random samples of the same size, from the same population, and fit the same OLS regression model repeatedly. Each random sample produces different estimates for the parameters in the regression equation. After this process, we can graph the distribution of estimates for each parameter. Statisticians refer to this type of distribution as a sampling distribution, which is a type of [probability distribution](https://statisticsbyjim.com/basics/probability-distributions/).

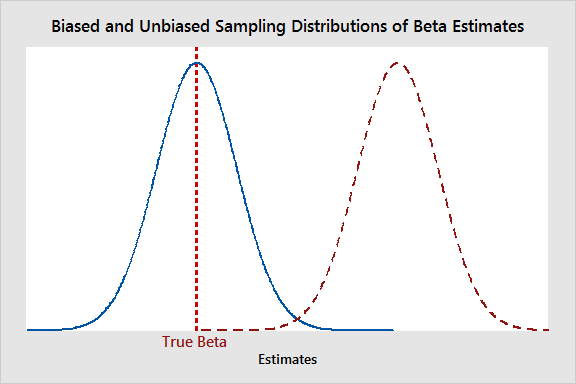
Keep in mind that each curve represents the sampling distribution of the estimates for a single parameter. The graphs below tell us which values of parameter estimates are more and less common. They also indicate how far estimates are likely to fall from the correct value.

Of course, when you conduct a real study, you’ll perform it once, not know the actual population value, and you definitely won’t see the sampling distribution. Instead, your analysis draws one value from the underlying sampling distribution for each parameter. However, using statistical principles, we can understand the properties of the sampling distributions without having to repeat a study many times. Isn’t the [field of statistics](https://statisticsbyjim.com/basics/importance-statistics/) grand?!

Hypothesis tests also use sampling distributions to calculate p-values and create confidence intervals. For more information about this process, read my post: [How Hypothesis Tests Work](https://statisticsbyjim.com/hypothesis-testing/hypothesis-tests-significance-levels-alpha-p-values/).

Unbiased Estimates: Sampling Distributions Centered on the True Population Parameter

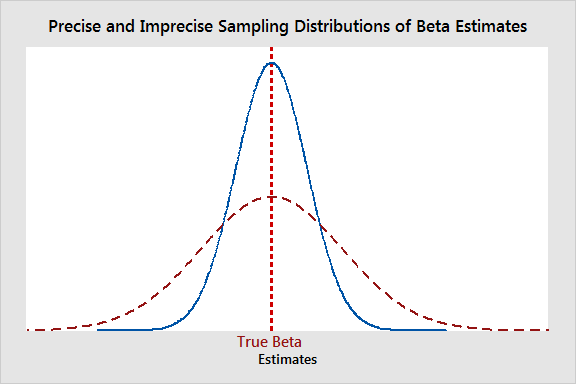
In the graph below, beta represents the true population value. The curve on the right centers on a value that is too high. This model tends to produce estimates that are too high, which is a positive bias. It is not correct on average. However, the curve on the left centers on the actual value of beta. That model produces parameter estimates that are correct on average. The expected value is the actual value of the population parameter. That’s what we want and satisfying the OLS assumptions helps us!



Keep in mind that the curve on the left doesn’t indicate that an individual study necessarily produces an estimate that is right on target. Instead, it means that OLS produces the correct estimate on average when the assumptions hold true. Different studies will generate values that are sometimes higher and sometimes lower—as opposed to having a tendency to be too high or too low.

Minimum Variance: Sampling Distributions are Tight Around the Population Parameter

In the graph below, both curves center on beta. However, one curve is wider than the other because the variances are different. Broader curves indicate that there is a higher probability that the estimates will be further away from the correct value. That’s not good. We want our estimates to be close to beta.



Both studies are correct on average. However, we want our estimates to follow the narrower curve because they’re likely to be closer to the correct value than the wider curve. The Gauss-Markov theorem states that satisfying the OLS assumptions keeps the sampling distribution as tight as possible for unbiased estimates.

The Best in BLUE refers to the sampling distribution with the minimum variance. That’s the tightest possible distribution of all unbiased linear estimation methods!

Gauss-Markov Theorem OLS Estimates and Sampling Distributions

As you can see, the best estimates are those that are unbiased and have the minimum variance. When your model satisfies the assumptions, the Gauss-Markov theorem states that the OLS procedure produces unbiased estimates that have the minimum variance. The sampling distributions are centered on the actual population value and are the tightest possible distributions. Finally, these aren’t just the best estimates that OLS can produce, but the best estimates that any linear model estimator can produce. Powerful stuff!

11. Describe two major issues with regression analysis.

### Problem 3

You are interested in knowing the relationship between the weather and tourism levels. To investigate, you collect data from the touristic centre in a city during one month in the summer, counting the number of people that arrive at the square at the same time every day. Given the data set below, what is the **correlation** between temperature and tourism? Interpret the correlation and name a few other reasons why these two variables are or are not related.

|  |  |
| --- | --- |
| **Temperature** | **Number of Visitors** |
| 12 | 87 |
| 21 | 150 |
| 20 | 110 |
| 25 | 90 |
| 17 | 85 |
| 15 | 70 |
| 13 | 90 |

### Problem 4

There are two variables that need to be studied: weight loss and days spent exercising one month. You are given a data set in which individuals have been asked the number of days they exercise for more than half an hour in one month. What kind of **regression model** can you use here? What are the results of this regression given the data set below. Interpret the model’s estimators.

|  |  |
| --- | --- |
| **Exercise Days** | **Weight Loss (in kg)** |
| 0 | 4 |
| 4 | 1 |
| 8 | 1.5 |
| 12 | 2 |
| 16 | 4 |
| 20 | 5 |
| 24 | 2 |

12. How can the linear regression model's accuracy be improved?

Mastering linear regression is an invaluable skill for data-driven decision-making. By following the tips outlined in this guide, you can enhance the quality and accuracy of your linear regression models. Remember to emphasize thorough data pre-processing, effective feature selection, and careful model evaluation. Implement regularization techniques, cross-validation, and hyper parameter tuning to fine-tune your models and achieve optimal performance. With these strategies in place, you can confidently utilize linear regression to uncover insights, make predictions, and drive business success.

13. Using an example, describe the polynomial regression model in detail.

## ****What is a Polynomial Regression?****

* There are some relationships that a researcher will hypothesize is curvilinear. Clearly, such types of cases will include a polynomial term.
* Inspection of residuals. If we try to fit a linear model to curved data, a scatter plot of residuals (Y-axis) on the predictor (X-axis) will have patches of many positive residuals in the middle. Hence in such a situation, it is not appropriate.
* An assumption in the usual multiple linear regression analysis is that all the independent variables are independent. In the [polynomial regression](https://www.geeksforgeeks.org/polynomial-regression-from-scratch-using-python/) model, this assumption is not satisfied.

## Why Polynomial Regression?

Polynomial regression is a type of regression analysis used in statistics and machine learning when the relationship between the independent variable (input) and the dependent variable (output) is not linear. While simple linear regression models the relationship as a straight line, polynomial regression allows for more flexibility by fitting a polynomial equation to the data.

When the relationship between the variables is better represented by a curve rather than a straight line, polynomial regression can capture the non-linear patterns in the data.

## How does a Polynomial Regression work?

If we observe closely then we will realize that to evolve from linear regression to polynomial regression. We are just supposed to add the higher-order terms of the dependent features in the feature space. This is sometimes also known as [feature engineering](https://www.geeksforgeeks.org/what-is-feature-engineering/) but not exactly.

When the relationship is non-linear, a polynomial regression model introduces higher-degree polynomial terms.

The general form of a polynomial regression equation of degree n is:

where,

* y is the dependent variable.
* x is the independent variable.
* ​ are the coefficients of the polynomial terms.
* n is the degree of the polynomial.
* represents the error term.

The basic goal of regression analysis is to model the expected value of a dependent variable y in terms of the value of an independent variable x. In simple [linear regression](https://www.geeksforgeeks.org/ml-linear-regression/), we used the following equation –

**y** = a + bx + e

Here y is a dependent variable, a is the y-intercept, b is the slope and e is the error rate. In many cases, this linear model will not work out For example if we analyze the production of chemical synthesis in terms of the temperature at which the synthesis takes place in such cases we use a quadratic model.

Here,

* y is the dependent variable on x
* a is the y-intercept and e is the error rate.

In general, we can model it for the nth value.

Since the regression function is linear in terms of unknown variables, hence these models are linear from the point of estimation. Hence through the [Least Square technique](https://www.geeksforgeeks.org/least-square-regression-line/), response value (y) can be computed.

By including higher-degree terms (quadratic, cubic, etc.), the model can capture the non-linear patterns in the data.

1. **The choice of the polynomial degree (*n*)** is a crucial aspect of polynomial regression. A higher degree allows the model to fit the training data more closely, but it may also lead to overfitting, especially if the degree is too high. Therefore, the degree should be chosen based on the complexity of the underlying relationship in the data.
2. The polynomial regression model is trained to**find the coefficients** that minimize the difference between the predicted values and the actual values in the training data.
3. Once the model is trained, it can be used to make predictions on new, unseen data. The polynomial equation captures the non-linear patterns observed in the training data, allowing the model to generalize to non-linear relationships.

## Polynomial Regression Real-Life Example

Let’s consider a real-life example to illustrate the application of polynomial regression. Suppose you are working in the field of finance, and you are analyzing the relationship between the years of experience (in years) an employee has and their corresponding salary (in dollars). You suspect that the relationship might not be linear and that higher degrees of the polynomial might better capture the salary progression over time.

| **Years of Experience** | **Salary (in dollars)** |
| --- | --- |
| **1** | 50,000 |
| **2** | 55,000 |
| **3** | 65,000 |
| **4** | 80,000 |
| **5** | 110,000 |
| **6** | 150,000 |
| **7** | 200,000 |

Now, let’s apply polynomial regression to model the relationship between years of experience and salary. We’ll use a quadratic polynomial (degree 2) for this example.

The quadratic polynomial regression equation is:

Salary= ×Experience+​×Experience^2+

Now, to find the coefficients that minimize the difference between the predicted salaries and the actual salaries in the dataset we can use a method of least squares. The objective is to minimize the sum of squared differences between the predicted values and the actual values.

## ****Polynomial Regression implementations using Python****

To get the Dataset used for the analysis of Polynomial Regression, click [here](https://media.geeksforgeeks.org/wp-content/uploads/data.csv). Import the important libraries and the dataset we are using to perform Polynomial Regression.

[Python](https://www.geeksforgeeks.org/python-programming-language/) libraries make it very easy for us to handle the data and perform typical and complex tasks with a single line of code.

* [**Pandas**](https://www.geeksforgeeks.org/pandas-tutorial/) – This library helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go.
* [**Numpy**](https://www.geeksforgeeks.org/numpy-tutorial/)– Numpy arrays are very fast and can perform large computations in a very short time.
* [**Matplotlib**](https://www.geeksforgeeks.org/matplotlib-tutorial/)/**[Seaborn](https://www.geeksforgeeks.org/python-seaborn-tutorial/)**– This library is used to draw visualizations.
* Sklearn – This module contains multiple libraries having pre-implemented functions to perform tasks from data preprocessing to model development and evaluation.

14. Provide a detailed explanation of logistic regression.

## What is Logistic Regression?

Logistic regression is used for binary [classification](https://www.geeksforgeeks.org/getting-started-with-classification/) where we use [sigmoid function](https://www.geeksforgeeks.org/derivative-of-the-sigmoid-function/), that takes input as independent variables and produces a probability value between 0 and 1.

For example, we have two classes Class 0 and Class 1 if the value of the logistic function for an input is greater than 0.5 (threshold value) then it belongs to Class 1 otherwise it belongs to Class 0. It’s referred to as regression because it is the extension of[linear regression](https://www.geeksforgeeks.org/ml-linear-regression/) but is mainly used for classification problems.

### Key Points:

* Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value.
* It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
* In Logistic regression, instead of fitting a regression line, we fit an “S” shaped logistic function, which predicts two maximum values (0 or 1).

## ****Logistic Function – Sigmoid Function****

* The sigmoid function is a mathematical function used to map the predicted values to probabilities.
* It maps any real value into another value within a range of 0 and 1. The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the “S” form.
* The S-form curve is called the Sigmoid function or the logistic function.
* In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

## Types of Logistic Regression

On the basis of the categories, Logistic Regression can be classified into three types:

1. **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
2. **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as “cat”, “dogs”, or “sheep”
3. **Ordinal:**In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as “low”, “Medium”, or “High”.

## Assumptions of Logistic Regression

We will explore the assumptions of logistic regression as understanding these assumptions is important to ensure that we are using appropriate application of the model. The assumption include:

1. Independent observations: Each observation is independent of the other. meaning there is no correlation between any input variables.
2. Binary dependent variables: It takes the assumption that the dependent variable must be binary or dichotomous, meaning it can take only two values. For more than two categories SoftMax functions are used.
3. Linearity relationship between independent variables and log odds: The relationship between the independent variables and the log odds of the dependent variable should be linear.
4. No outliers: There should be no outliers in the dataset.
5. Large sample size: The sample size is sufficiently large

## ****Terminologies involved in Logistic Regression****

Here are some common terms involved in logistic regression:

* **Independent variables:** The input characteristics or predictor factors applied to the dependent variable’s predictions.
* **Dependent variable:** The target variable in a logistic regression model, which we are trying to predict.
* **Logistic function:** The formula used to represent how the independent and dependent variables relate to one another. The logistic function transforms the input variables into a probability value between 0 and 1, which represents the likelihood of the dependent variable being 1 or 0.
* **Odds:**It is the ratio of something occurring to something not occurring. it is different from probability as the probability is the ratio of something occurring to everything that could possibly occur.
* **Log-odds:**The log-odds, also known as the logit function, is the natural logarithm of the odds. In logistic regression, the log odds of the dependent variable are modeled as a linear combination of the independent variables and the intercept.
* **Coefficient:**The logistic regression model’s estimated parameters, show how the independent and dependent variables relate to one another.
* **Intercept:**A constant term in the logistic regression model, which represents the log odds when all independent variables are equal to zero.
* [**Maximum likelihood estimation**](https://www.geeksforgeeks.org/probability-density-estimation-maximum-likelihood-estimation/)**:** The method used to estimate the coefficients of the logistic regression model, which maximizes the likelihood of observing the data given the model.

## How does Logistic Regression work?

The logistic regression model transforms the [linear regression](https://www.geeksforgeeks.org/ml-linear-regression/) function continuous value output into categorical value output using a sigmoid function, which maps any real-valued set of independent variables input into a value between 0 and 1. This function is known as the logistic function.

Let the independent input features be:

 𝑋=[𝑥11 …𝑥1𝑚𝑥21 …𝑥2𝑚 ⋮⋱ ⋮ 𝑥𝑛1 …𝑥𝑛𝑚]*X*=​*x*11​ *x*21​  ⋮*xn*1​ ​……⋱ …​*x*1*m*​*x*2*m*​⋮ *xnm*​​​

 and the dependent variable is Y having only binary value i.e. 0 or 1.

𝑌={0 if 𝐶𝑙𝑎𝑠𝑠11 if 𝐶𝑙𝑎𝑠𝑠2*Y*={01​ if *Class*1 if *Class*2​

then, apply the multi-linear function to the input variables X.

𝑧=(∑𝑖=1𝑛𝑤𝑖𝑥𝑖)+𝑏*z*=(∑*i*=1*n*​*wi*​*xi*​)+*b*

Here 𝑥𝑖*xi*​ is the ith observation of X, 𝑤𝑖=[𝑤1,𝑤2,𝑤3,⋯,𝑤𝑚]*wi*​=[*w*1​,*w*2​,*w*3​,⋯,*wm*​] is the weights or Coefficient, and b is the bias term also known as intercept. simply this can be represented as the dot product of weight and bias.

𝑧=𝑤⋅𝑋+𝑏*z*=*w*⋅*X*+*b*

whatever we discussed above is the [linear regression](https://www.geeksforgeeks.org/ml-linear-regression/).

### Sigmoid Function

Now we use the [sigmoid function](https://www.geeksforgeeks.org/derivative-of-the-sigmoid-function/) where the input will be z and we find the probability between 0 and 1. i.e. predicted y.

(𝑧)=11+𝑒−𝑧*σ*(*z*)=1+*e*−*z*1​



*Sigmoid function*

As shown above, the figure sigmoid function converts the continuous variable data into the [probability](https://www.geeksforgeeks.org/probability-gq/) i.e. between 0 and 1.

* 𝜎(𝑧)   *σ*(*z*)  tends towards 1 as 𝑧→∞*z*→∞
* 𝜎(𝑧)   *σ*(*z*)  tends towards 0 as 𝑧→−∞*z*→−∞
* 𝜎(𝑧)   *σ*(*z*)  is always bounded between 0 and 1

where the probability of being a class can be measured as:

𝑃(𝑦=1)=𝜎(𝑧)𝑃(𝑦=0)=1−𝜎(𝑧)*P*(*y*=1)=*σ*(*z*)*P*(*y*=0)=1−*σ*(*z*)

## Code Implementation for Logistic Regression

### ****Binomial Logistic regression:****

Target variable can have only 2 possible types: “0” or “1” which may represent “win” vs “loss”, “pass” vs “fail”, “dead” vs “alive”, etc., in this case, sigmoid functions are used, which is already discussed above.

Importing necessary libraries based on the requirement of model. This Python code shows how to use the breast cancer dataset to implement a Logistic Regression model for classification.

15. What are the logistic regression assumptions?

# Assumptions of Logistic Regression

[Logistic regression](https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-logistic-regression/) does not make many of the key assumptions of [linear regression](https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/)and [general linear models](https://www.statisticssolutions.com/academic-solutions/resources/directory-of-statistical-analyses/generalized-linear-models/) that are based on ordinary least squares algorithms – particularly regarding linearity, normality, [homoscedasticity](https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/homoscedasticity/), and measurement level.

First, logistic regression does not require a linear relationship between the dependent and independent variables.  Second, the error terms (residuals) do not need to be normally distributed.  Third, homoscedasticity is not required.  Finally, the dependent variable in logistic regression is not measured on an interval or ratio scale.  
However, some other assumptions still apply.

First, binary logistic regression requires the dependent variable to be binary and ordinal logistic regression requires the dependent variable to be ordinal.

Second, logistic regression requires the observations to be independent of each other.  In other words, the observations should not come from repeated measurements or matched data.

Third, logistic regression requires there to be little or no [multicollinearity](https://www.statisticssolutions.com/multicollinearity/) among the independent variables.  This means that the independent variables should not be too highly correlated with each other.

Fourth, logistic regression assumes linearity of independent variables and log odds of the dependent variable. Although this analysis does not require the dependent and independent variables to be related linearly, it requires that the independent variables are linearly related to the log odds of the dependent variable.

Finally, logistic regression typically requires a large sample size.  A general guideline is that you need at minimum of 10 cases with the least frequent outcome for each independent variable in your model. For example, if you have 5 independent variables and the expected probability of your least frequent outcome is .10, then you would need a minimum sample size of 500 (10\*5 / .10).

16. Go through the details of maximum likelihood estimation.

*This article is about the statistical techniques. For computer data storage, see*[*partial-response maximum-likelihood*](https://en.wikipedia.org/wiki/Partial-response_maximum-likelihood)*.*

In [statistics](https://en.wikipedia.org/wiki/Statistics), **maximum likelihood estimation** (**MLE**) is a method of [estimating](https://en.wikipedia.org/wiki/Estimation_theory) the [parameters](https://en.wikipedia.org/wiki/Statistical_parameter) of an assumed [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution), given some observed data. This is achieved by [maximizing](https://en.wikipedia.org/wiki/Mathematical_optimization) a [likelihood function](https://en.wikipedia.org/wiki/Likelihood_function) so that, under the assumed [statistical model](https://en.wikipedia.org/wiki/Statistical_model), the [observed data](https://en.wikipedia.org/wiki/Realization_(probability)) is most probable. The [point](https://en.wikipedia.org/wiki/Point_estimate) in the [parameter space](https://en.wikipedia.org/wiki/Parameter_space) that maximizes the likelihood function is called the maximum likelihood estimate.[[1]](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation#cite_note-1) The logic of maximum likelihood is both intuitive and flexible, and as such the method has become a dominant means of [statistical inference](https://en.wikipedia.org/wiki/Statistical_inference).[[2]](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation#cite_note-2)[[3]](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation#cite_note-3)[[4]](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation#cite_note-4)

If the likelihood function is [differentiable](https://en.wikipedia.org/wiki/Differentiable_function), the [derivative test](https://en.wikipedia.org/wiki/Derivative_test) for finding maxima can be applied. In some cases, the first-order conditions of the likelihood function can be solved analytically; for instance, the [ordinary least squares](https://en.wikipedia.org/wiki/Ordinary_least_squares) estimator for a [linear regression](https://en.wikipedia.org/wiki/Linear_regression) model maximizes the likelihood when the random errors are assumed to have [normal](https://en.wikipedia.org/wiki/Normal_distribution) distributions with the same variance.[[5]](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation#cite_note-5)

From the perspective of [Bayesian inference](https://en.wikipedia.org/wiki/Bayesian_inference), MLE is generally equivalent to [maximum a posteriori (MAP) estimation](https://en.wikipedia.org/wiki/Maximum_a_posteriori_estimation) with [uniform](https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)) [prior distributions](https://en.wikipedia.org/wiki/Prior_probability) (or a [normal](https://en.wikipedia.org/wiki/Normal_distribution) prior distribution with a standard deviation of infinity). In [frequentist inference](https://en.wikipedia.org/wiki/Frequentist_inference" \o "Frequentist inference), MLE is a special case of an [extremum estimator](https://en.wikipedia.org/wiki/Extremum_estimator" \o "Extremum estimator), with the objective function being the likelihood.