Linear Regression

Linear Regression

Linear Regression is a statistical technique that establishes an equation that allows the unknown value of one variable to be estimated from the known value of one other variable is simple linear regression analysis.

Linear Regression is used to measure the relationship between one or more *predictor variable(s)* and one *outcome variable*. Linear regression is commonly used for predictive analysis and modeling. For example, it can be used to quantify the relative impacts of age, gender, and diet (the predictor variables) on height (the outcome variable).

Regression models describe the relationship between variables by fitting a line to the observed data. Linear regression models use a straight line, while logistic and nonlinear regression models use a curved line. Regression allows you to estimate how a ***dependent*** variable (y) changes as the ***independent*** (*x*) variable change.

Linear Regression Basics

A linear regression is a statistical model that attempts to show the relationship between two variables with *a linear equation*. A regression analysis involves graphing a line over a set of data points that most closely fits the overall shape of the data.

Simple linear regression is used to estimate the relationship between **two quantitative variables.**

You can use simple linear regression when you want to know:

1. How strong the relationship is between two variables, example: the relationship between incomes and expenses is very strong and the correlation ***r*** is close to +1.

   Example of independent variable (x) and dependent variable (y):

Variable X	Variable Y	Correlation (r)
Salary	Taxes paid	Positive
Shyness	Number of people greeted at party	Negative
Price of car	Prestige of car	Positive
Price of tennis shoe	Foot support	Zero
Time of use of flashlight	Battery life	Negative
Weight in lbs.	Average daily caloric intake	Positive
Price of quartz watch	Accuracy of time kept	Zero
Salary of sales people	Number of cars sold	Positive

2. You want to know the value of the dependent variable (*y*) at a certain value of the independent variable (*x*). Example: the amount of sales at a certain level of advertising expenditure.

Linear Regression in Business

Linear regressions can be used in business to evaluate trends and make estimates or forecasts. For example, if a company's sales have increased steadily every month for the past few years, conducting a linear analysis on the sales data with monthly sales on the y-axis and time on the x-axis would produce a line that that depicts the upward trend in sales. After creating the trend line, the company could use the slope of the line to forecast sales in future months.

The major uses for regression in business analysis are:
(1) identifying the strength of predictors,
(2) forecasting an effect, and
(3) trend forecasting.

First, the regression might be used to ***identify the strength of the effect*** that the independent variable(s) have on a dependent variable. Typical questions are: "What is the strength of relationship between dose and effect, sales and marketing spending, or age and income?"
Second, it can be used to ***forecast effects or impact of changes***. That is, the regression analysis helps us to understand how much the dependent variable changes with a change in one or more independent variables. A typical question is, "how much additional sales income do I get for each additional $1, 000 spent on marketing?"

Third, regression analysis ***predicts trends and future values***. The regression analysis can be used to get point estimates. A typical question is, "what will the price of gold be in 6 months?"

Regression Analysis

Regression Analysis

In statistics, regression analysis is a mathematical method used to understand the relationship between a dependent variable and an independent variable. Results of this analysis demonstrate the strength of the relationship between the two variables and if the dependent variable is significantly impacted by the independent variable.

Least Square Linear Regression

In a cause and effect relationship, the **independent variable** is the **cause**, and the **dependent variable** is the **effect**. **Least squares linear regression** is a method for predicting the value of a dependent variable *Y*, based on the value of an independent variable *X*.

Regression Analysis- Simple Linear Regression

Regression analysis is used to find equations that fit data. Once we have the regression equation, we can use the model to make predictions. One type of regression analysis is linear analysis. When a **correlation coefficient** shows that data is likely to be able to predict future outcomes and **a scatter plot** of the data appears to form ***a straight line***, you can use simple linear regression to find a predictive function. Simple linear regression is a model that assess the relationship between a dependent variable and an independent variable.

Least Square Regression Line

Linear regression finds the straight line, called the **least squares regression line** or LSRL, that best represents observations in two variables data set. Suppose Y is a dependent variable, and X is an independent variable.

The population regression line is:

$$Y = B_0 + B_1X$$

where: B_0 is a constant,
B_1 is the regression coefficient,
X is the value of the independent variable, and
Y is the value of the dependent variable.

Estimated Least Square Regression Line

A regression line calculated from sample data by the method of least squares is called **an estimated least-squares regression line** or **sample regression line**. In its equation, the values of **a** and **b** are the estimated regression coefficients.

Given a random sample of observations, the population regression line is estimated by sample linear regression as follows.

The Estimated Least-Squares Regression Line or
Sample Regression Line or Sample Linear Regression:

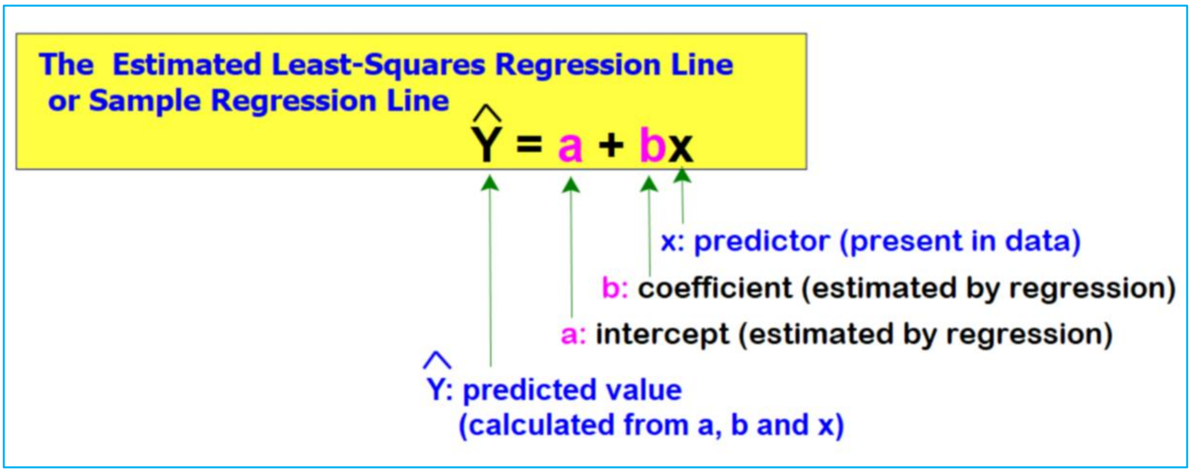
$$\hat{y} = b_0 + b_1x$$

or

$$\hat{y} = a + bx$$

where:

b_0	or a	is an *intercept* (constant),
b_1	or b	is the *slope* or regression coefficient,
	x	is the value of the independent variable, and
	$\hat{y}$	is the *predicted* value of the dependent variable.



The Estimated Least-Squares Regression Line or Sample Regression Line:

$$\hat{Y} = a + bx$$

Least square linear regression equation

$\hat{Y}_c = a + bX$

a is the sample regression constant, or Y -intercept

b is the sample regression coefficient, or slope of the line

n is the sample size

$b = \dfrac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2}$	$a = \bar{Y} - b\bar{X}$

Example 1 The data set below to represent monthly sales calls, and a corresponding number of deals closed over 12 months in year 2020 period.

Sale period	Sales calls (x)	Deals Closed (y)
January	20	50
February	20	30
March	40	60
April	10	30
May	30	70
June	10	40
July	30	60
August	20	40
September	30	60
October	10	40
November	20	30
December	30	70

a) A scatter diagram to determine the relationship between the Sales calls and number of deals closed.
b) Find an Estimated Least- Square regression line or sample regression line;
c) Graph this Least- Square regression line in a scatter diagram.
d) Estimate the number of deals closed if the number of sales calls are 50.
Find:

Solution

Sale period	Sales calls (x)	Deals Closed (y)	xy	x^2
January	20	50		
February	20	30		
March	40	60		
April	10	30		
May	30	70		
June	10	40		
July	30	60		
August	20	40		
September	30	60		
October	10	40		
November	20	30		
December	30	70		
n =	$\sum x =$	$\sum x =$	$\sum xy =$	$\sum x^2 =$

Formula: $a = \overline{Y} - b\overline{X}$

$$\overline{Y} = \frac{\sum Y}{n}$$

$$\overline{X} = \frac{\sum \overline{X}}{n}$$

a = 20.89431

$$b = \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sum X^2 - n\overline{X}^2}$$

b = 1.21951

an Estimated Least- Square regression line or sample regression line is

$$\hat{Y}_x = a + bX$$

$\hat{Y}_x$ =

d) Estimate the number of deals closed if the number of sales calls are 50.

x = 50 $\rightarrow$ $\hat{y}_{50}$ = a + b(50)

➡ $\hat{y}_{50}$ =