Introduction to Regression Analysis

- Regression analysis is the most often applied technique of statistical analysis and modeling.
- If two variables are involved, the variable that is the basis of the estimation, is conventionally called the independent variable and the variable whose value is to be estimated+ is called the dependent variable.
- In general, it is used to model a response variable (Y) as a function of one or more driver variables (X_1, X_2, ..., X_p).
- The functional form used is:

  $$Y_i = \square_0 + \square_1 X_{1i} + \square_2 X_{2i} + ... + \square_p X_{pi} + \square$$

- The dependent variable is variously known as explained variables, predictand, response and endogenous variables.
- While the independent variable is known as explanatory, regressor and exogenous variable

The term ε in the model is referred to as a "random error term" and may reflect a number of things including the general idea that knowledge of the driver variables will not ordinarily lead to perfect reconstruction of the response.

- If there is only one driver variable, X, then we usually speak of "simple" linear regression analysis.
- When the model involves
  - (a) multiple driver variables,
  - (b) a driver variable in multiple forms, or
  - (c) a mixture of these, the we speak of "multiple linear regression analysis".
- The "linear" portion of the terminology refers to the response variable being expressed as a "linear combination" of the driver variables.

In regression analysis, the data used to describe the relationship between the variables are primarily measured on interval scale. the chief advantage of using the interval level of measurement is that, with such data it is possible to describe the relationship between variables more exactly employing mathematical equation. This in turn allows more accurate prediction of one variable from the knowledge of the other variables, which is one of the most important objectives of regression analysis.

It is important to note that if the relationship between X and Y is curvilinear , the regression line will be a curved line rather than straight line. The greater the strength of relationships between X and Y the better is the prediction

The problem is presented to the mathematician as follows: "The values of a and b in the linear model $Y'_i = a + b X_i$ are to be found which minimize the algebraic expression ."

The mathematician begins as follows:

The result becomes:

Using a similar procedure to find the value of a yields:

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

	Y_i	X_i^2	X_iY_i	
	13	23	169	299
	20	18	400	360
	10	35	100	350
	33	10	1089	330
	15	27	225	405
SUM	91	113	1983	1744

$$N = 5$$
$$\sum X = 91$$
$$\sum Y = 113$$
$$\sum X^2 = 1983$$
$$\sum XY = 1744$$

$$b=\frac{N\sum XY-\sum X\sum Y}{N\sum X^2-(\sum X)^2}$$

$$b=\frac{5*1744 - (91*113)}{5*1983 - 91^2}$$

$$b=\frac{8720 - 10283}{9915-8281}$$

$$b=\frac{-1563}{1634} = -.9565$$

$$\bar{X}=18.2$$
$$\bar{Y}=22.6$$
$$b=-.957$$
$$a=\bar{Y}-b\bar{X}$$
$$a=22.6-(-.957*18.2)$$
$$a=40.01$$

THE REGRESSION MODEL

The situation using the regression model is analogous to that of the interviewers, except instead of using interviewers, predictions are made by performing a linear transformation of the predictor variable. Rather than interviewers in the above example, the predicted value would be obtained by a linear transformation of the score. The prediction takes the form

$$Y' = a + bX$$

where a and b are parameters in the regression model

Types of regression analysis:

Regression analysis is generally classified into two kinds: simple and multiple. Simple
regression involves only two variables, one of which is dependent variable and the other
Is explanatory(independent) variable. The associated model in the case of simple regression will be a simple regression model.

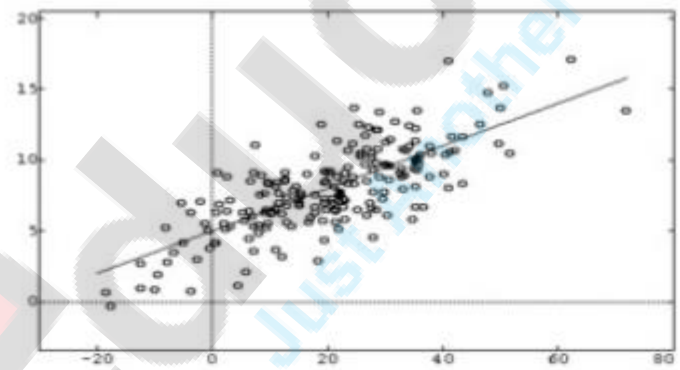
- A regression analysis may involve a linear model or a nonlinear model.

The term linear can be interpreted in two different ways:

1. Linear in variable
2. Linearity in the parameter

LINEAR REGRESSION

In linear regression, the model specification is that the dependent variable, y_i is a linear combination of the *parameters* (but need not be linear in the *independent variables*). For example, in simple linear regression for modeling n data points there is one independent variable: x_i, and two parameters, β_0 and β_1:



In the case of simple regression, the formulas for the least squares estimates are

$$Y = a + bX$$

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \qquad a = \frac{\sum Y - b\sum X}{N}$$

Where,

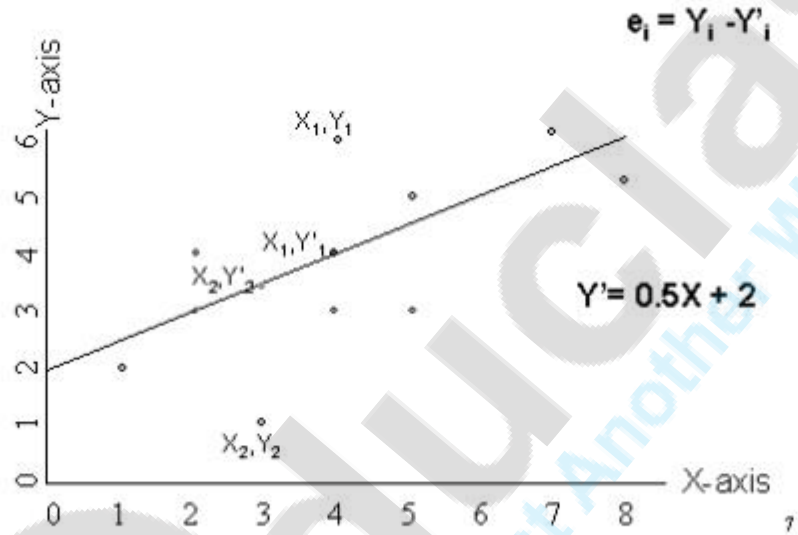
N = number of observations, or years

X = a year index (decade)

Y = population size for given census years

Scatter plot

- regression



Uses of Regression Analysis

1. Regression analysis helps in establishing a functional Relationship between two or more variables.

2. Since most of the problems of economic analysis are based on cause and effect relationships, the regression analysis is a highly valuable tool in economic and business research.

3. Regression analysis predicts the values of dependent variables from the values of independent variables.

4. We can calculate coefficient of correlation (r) and coefficient of determination (R^2) with the help of regression coefficients.