Frequently Asked Questions

1. What do you mean by regression?
 Answer:
 The dictionary meaning of the term regression is the act of returning or going back. It is usually used to see the relationship between the two variables.

2. Who used the term regression for the first time?
 Answer:
 The term regression was first used by Francis Galton towards the end of nineteenth century. This term was introduced by him in the paper 'Regression towards Mediocrity in Hereditary Stature'.

3. Define Regression analysis.
 Answer:
 Regression analysis is a branch of statistical theory that is widely used in almost all the scientific disciplines. According to M. M. Blair, "Regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data".

4. What do you call the regression study of two variables?
 Answer:
 The regression study which confines itself to a study of only two variables is called simple regression.

5. What do you call the regression study of more than two variables?
 Answer:
 The regression analysis which studies more than two variables at a time is called multiple regression.

6. Write the utility of regression analysis.
 Answer:
 In the field of business this tool of statistical analysis is very widely used. Businessmen are interested in predicting future production, consumption, investment, prices, profits, sales etc. In fact the success of a businessman depends on the correctness of the various estimates that he is required to make.
 In sociological studies and in the field of economic planning, projections of population, birth rates, death rates and other similar variables are of great use.
 In our day to day life we come across many variables which are inter-related. For example, with a rise in price, the demand of a commodity goes down, or with better monsoons the output of agricultural product increase or the effect of expenditure on publicity may lead to rise in the volume of sales. With the help of regression analysis we can estimate or predict the effect of one variable on the other.

7. What are different methods of studying regression?
 Answer:
 Regression can be studied either graphically or algebraically.

8. Explain graphical study of regression.
 Answer:
         When regression is studied with the help of graphic methods, we have to draw a

scatter diagram. A scatter diagram contains one point for one pair of values of X and Y

variable. Usually, X variable is shown on the horizontal scale and Y variable on the vertical scale. When all related pair of values have been plotted on a scatter diagram, we have to draw 2 regression lines to predict the values of X and Y variables. The regression line which is used to predict the values of Y for a value of X is called the Regression line of Y on X. Similarly the regression line which is used to predict a value of X for a value of Y is called regression line of X on Y. If the coefficient of correlation between X and Y is perfect, that is its value either +1 or -1, there will be only one regression line as the variations in the two series is such cases always increase or decrease by a constant figure. In other words we say that the 2 regression lines will be identical if the correlation between the 2 variables is perfect.

9. Differentiate between correlation and regression.
Answer:

Sl. No.	Correlation	Regression
1	It measures the relationship between two variables which vary in the same or opposite directions.	It means going back or act of return. It is a mathematical measure which shows the average relationship between two variables.
2	Here, both X and Y variables are random variables.	Here X is random variable and Y is a fixed variable. However, both variables may be random variables.
3	Correlation tells the degree of relationship between the two variables and not the cause and effect of the variables	Regression points out the cause and effect of relationship between the variables. It establishes a functional relationship
4.	Correlation is confined only to linear relationship between two variables.	Regression studies both linear and non-linear relationship between the variables.
5	It is immaterial whether X variable depends on Y and Y depends on X variable.	There is a functional relationship between the two variables.
6	There can be nonsense or spurious correlation between the two variables.	There is no such non-sense regression equation.
7.	Coefficient of correlation is not affected by change of scale or change of origin.	Regression coefficient is independent of change of origin by not of change of scale.
8.	Here, both the variables must be positively associated or vice-versa.	It explains that the decrease in one variable increase in the other variable.

10. What is the importance of regression analysis?
Answer:
By the study of regression analysis we are able to obtain the most probable values of one variable from the known values of another variable. For example, if two series relating to price and supply are correlated, we can find out what would be the effect on price if the supply of the commodity has increased or decreased to a particular level. The regression can also be used in natural, physical, and social science.

11. What do you mean by error in regression?

Answer:
The theory formulates exact functional relationship among the variables. But dealing with common data even an ordinary investigator will feel that all observations do not fall exactly on a straight line or any other the smooth functions. The best we can expect is that the observed quantities will be closer to the line, that is, why our regression model requires extension and stochastic disturbance term. The introduced term is known as disturbance or error term, because it represents the effects of all those factors which are not suspected by the investigator.

12. What are the different assumptions made under regression model?
 Answer:
   1. Homogeneous variance: The distribution of the dependent variable has the same spread. Formally this means that the probability distribution has the same variance sigma square
   2. Linearity: The mean of the dependent variable is a straight line, known as the true population regression line.
   3. Independence: The random variables are statistically independent.

13. Why errors exist in the regression?
 Answer:
 The error exist because of
   1. Measurement error: There are various reasons why the dependent variable is not measured correctly. For example, in a study of consumption of families at various income levels, the measurement error in consumption might consist of budget and reporting inaccuracies.
   2. Inherent Variability: Even if there were no measurement error, repetition of an experiment using exactly same amount of income would result in different levels of consumption.

14. Explain RMS error of regression.
 Answer:
 The regression line does not pass through all the points on the scatter plot exactly unless the correlation coefficient is ±1. In general, the data are scattered around the regression line. Each datum will have a vertical residual from the regression line; the sizes of the vertical residuals will vary from datum to datum. The rms of the vertical residuals measures the typical vertical distance of a datum from the regression line. The rms is a measure of the typical size of the elements in a list. Thus the rms of the vertical residuals is a measure of the typical vertical distance from the data to the regression line, which is the typical error in estimating the value of Y by the height of the regression line. A bit of algebra shows that the rms of the vertical residuals from the regression line or the rms error of regression is:
   $(1 - r^2)^{1/2} \times SD_Y.$

15. What is the range for RMS error of regression?
 Answer:
 The root mean square error of regression is always between 0 and SD_Y. It is zero when $r = \pm 1$ and SD_Y when $r = 0$.