Frequently Asked Questions:

1. What is regression analysis? Explain

   Ans. Regression analysis has been described as a study of the relationships between one variable, the response variable, and one or more other variables the predictor variables. Regression analysis addresses how the predictor variables influence, describe, or control the response

2. What is the goal of regression analysis?

   The goal of regression analysis is to express the response variable as a function of predictive variables. Identify which predictor variables are significant or most affect the response variable. Once such an expression is obtained, utilize the relationship to predict values of the response variable.

3. What is distribution of OLS estimator of regression coefficient β?

   Ans. Under the basic assumptions, the OLS estimator is distributed as multivariate normal with mean vector β and variance covariance matrix $(X'X)^{-1}$ σ^2.

4. What is the distribution of estimate of error variance σ^2?

   The estimate of error variance is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{\sigma^2}, \textit{ and is distributed } \chi^2 \textit{ with } (n-2)$$ degrees of freedom.

5. How to test the significance of regressor in the regression model?

   Ans. We test the hypothesis H_0: β=0, which means there is no relationship between Y_i and X_i or the regressor is insignificant. Using t-test

$$t_{obs} = \frac{\hat{\beta}}{\sqrt{\left(\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2}\right)}} = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

   If $|t_{obs}| > t_{\alpha/2,(n-2)}$, then reject H_0 at α% significance level.

6. What is prediction?

One of the major objective of regression analysis is prediction. Prediction is forecasting the value of the response variable for specified values of the regressors.

7. Define best linear unbiased predictor(BLUP).

Let $\hat{Y}_0$ be a linear function of sample observations $Y_1, Y_2, \ldots, Y_n$. $\hat{Y}_0$ is said to be best linear unbiased predictor of Y at $X=X_0$, if it satisfies the following conditions

(i) $E(\hat{Y}_0) = E(Y_0/X_0)$ and

(ii) $V(\hat{Y}_0)$ is minimum.

8. Give the expression for BLUP of regressand for a given value of regressor $X=X_0$.

For the linear regression model satisfying the basic ideal conditions, the BLUP of Y_0 is given by

$\hat{Y}_0 = \hat{\alpha}_{OLS} + \hat{\beta}_{OLS} X_0, \quad where\ \hat{\alpha} \quad and\ \hat{\beta}_{OLS}$ are OLS estimators of α and β.

9. What is the standard error of the fitted regression model?

Ans. The standard error of fitted regression model

Is square root of $V(\hat{Y}_0)$ where,

$$V(\hat{Y}_0) = \sigma^2\left(\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum x_i^2}\right)$$

10. Give 95% confidence interval for the mean response.

Ans. 95% confidence interval for the mean response for every value of X_0 is

$$\hat{Y}_0 \pm t_{0.025,n-2}\left[s\left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum x_i^2}\right)^{1/2}\right]$$

11. Define multiple linear regression model.

Ans. A multiple linear regression model corresponding to the i^{th} observation has the representation

$$Y_i = \sum_{j=1}^{k} \beta_j X_{ji}, \quad i = 1,2,...,n \qquad with\ X_{1i} = 1 \forall i = 1,..n$$

In matrix notation we can write the above equation for all the n observations as $Y = X\beta + \varepsilon$

12. State the basic assumptions on the multiple linear regression model.

We make the following assumptions on the multiple linear regression model usually called as **Basic ideal Conditions.**

A-1) X is non-stochastic. That is the regressors are not random variables.

A-2) Rank(X) =k. No multicollinearity problem.

A-3) E(ϵ)=0, No specification error

A-4) D(ϵ) =V=σ^2I , where I is the nxn identity matrix. This assumptions indicates absence of autocorrelation and heteroscedasticity)

A-5) ϵ is multivariate normal random vector with mean vector zero and variance-covariance matrix V=$\sigma^2 I_n$

A-6) $\lim_{n\to\infty}\left(\frac{X'X}{n}\right) = Q$ a finite and non-singular matrix. That is the model do not contain too big or too small regressors such as X_t =t, or $X_t = \lambda^t$, $0<\lambda<1$, t= 1,2,...

13. Define adjusted coefficient of determination and explain its significance.

Ans. In case of multiple linear regression model , as we increase the number of regressors by including redundant and irrelevant variables, there is a possibility of increase in R^2, misguiding overall goodness of fit of the model. Hence a supplementary measure $\overline{R}^2$ (or adjusted R^2) is defined as

$$\overline{R}^2 = 1 - \frac{(n-1)}{(n-k)}(1 - R^2)$$

Adjusted R^2 cannot be increased by simply increasing k.

14. Define OLS estimator of β.

Ans. The OLS estimator of β is given by

$$\hat{\beta} = (X'X)^{-1}X'Y$$

15. Can we use the test procedure discussed above for testing various hypothesis if the regression model violates the any of the basic assumptions.

    Ans. No. If the basic assumptions are violated OLS estimator and also test procedures based on OLS estimator are misleading and hence are invalied.