

# Sample Questions

Computer Engineering

**Subject Name:** Quantitative Analysis

**Semester:** VI

## Multiple Choice Questions

1.	<p>Q:Multiple linear regression (MLR) is a _____ type of statistical analysis.</p> <p>Options</p> <ul style="list-style-type: none"><li>a) Univariate</li><li>b) Bivariate</li><li>c) Multivariate</li><li>d) Trivariate</li></ul>
2.	<p>Q:A linear regression (LR) analysis produces the equation <math>Y = 0.4X + 3</math>. This indicates that:</p> <p>Options</p> <ul style="list-style-type: none"><li>When <math>Y = 0.4</math>, <math>X = 3</math></li><li>When <math>Y = 0</math>, <math>X = 3</math></li><li>When <math>X = 3</math>, <math>Y = 0.4</math></li><li>When <math>X = 0</math>, <math>Y = 3</math></li></ul>
3.	<p>Q:A LR analysis produces the equation <math>Y = -3.2X + 7</math>. This indicates that:</p> <p>Options</p> <ul style="list-style-type: none"><li>a) A 1 unit increase in X results in a 3.2 unit decrease in Y.</li><li>b) A 1 unit decrease in X results in a 3.2 unit decrease in Y.</li><li>c) A 1 unit increase in X results in a 3.2 unit increase in Y.</li><li>d) An X value of 0 would would increase Y by 7</li></ul>
4.	<p>When writing regression formulae, which of the following refers to the predicted value on the dependent variable (DV)?</p> <ul style="list-style-type: none"><li>a)Y</li><li>b)Y (hat)</li><li>c)X</li><li>d)X (hat)</li></ul>

5.	<p>In MLR, the square of the multiple correlation coefficient or <math>R^2</math> is called the</p> <ul style="list-style-type: none"> <li>a) Coefficient of determination</li> <li>b) Variance</li> <li>c) Covariance</li> <li>d) Cross-product</li> </ul>
6.	<p>Which of the following is true about the adjusted <math>R^2</math>?</p> <ul style="list-style-type: none"> <li>It is usually larger than the <math>R^2</math></li> <li>It is only used when there is just one predictor</li> <li>It is usually smaller than the <math>R^2</math></li> <li>It is used to determine whether residuals are normally distributed</li> </ul>
7.	<p>Least square method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the _____ deviations.</p> <ul style="list-style-type: none"> <li>a) Vertical</li> <li>b) Horizontal</li> <li>c) Both of these</li> <li>d) None of these</li> </ul>
8.	<p>A residual is defined as</p> <ul style="list-style-type: none"> <li>a) The difference between the actual Y values and the mean of Y.</li> <li>b) The difference between the actual Y values and the predicted Y values.</li> <li>c) The predicted value of Y for the average X value.</li> <li>d) The square root of the slope.</li> </ul>
9.	<p>The correct relationship between SST, SSR, and SSE is given by;</p> <ul style="list-style-type: none"> <li>a) <math>SSR = SST + SSE</math></li> <li>b) <math>SST = SSR + SSE</math></li> <li>c) <math>SSE = SSR - SST</math></li> <li>d) all of the above</li> </ul>

10.	<p>Below you are given a summary of the output from a simple linear regression analysis from a sample of size 15, SSR=100, SST = 152. The coefficient of determination is;</p> <p>a) 0.5200  b) 0.6579  c) 0.8111  d) 1.52</p>
11.	<p>Significance for the coefficients (b) is determined by</p> <p>a)an F-test.  b)an R<sup>2</sup> test.  c)a correlation coefficient.  d)a t-test.</p>
12.	<p>Q:A researcher polls people as they walk by on the street.</p> <p>Options</p> <p>a) Systematic Random Sample  b) Convenience Sampling  c) Judgmental Sampling  d) Quota Sampling</p>
13.	<p>Q:Inspectors for a hospital chain with multiple locations randomly select some of their locations for a cleanliness check of their operating rooms.</p> <p>Options</p> <p>a) Cluster sampling  b) Stratified Sampling  c) Quota Sampling  d) Snowball Sampling</p>
14.	<p>Q: The runs scored by a batsman in 5 ODIs are 31,97,112, 63, and 12. The standard deviation is</p> <p>Options</p> <p>1: 24.79</p>

	<p>2: 23.79 3: 25.79 4: 26.79</p>
15.	<p>Q: Find the mode of the call received on 7 consecutive day 11,13,13,17,19,23,25</p> <p>Options</p> <p>1: 11 2: 13 3: 17 4: 23</p>
16.	<p>Q: Find the median of the call received on 7 consecutive days 11,13, 17, 13, 23,25,19</p> <p>Options</p> <p>1: 13 2: 23 3: 25 4: 17</p>
17.	<p>Q: If the probability of hitting an object is 0.8, find the variance</p> <p>Options</p> <p>1: 0.18 2: 0.16 3: 0.14 4: 0.12</p>
18.	<p>Q: <math>E(X) = \lambda</math> is used for which distribution?</p> <p>Options</p> <p>1: Binomial distribution 2: Poisson's distribution 3: Bernoulli's distribution 4: Laplace distribution</p>
19.	<p>Q: The classification of data on geographical basis is also called as</p> <p>Options</p> <p>1: reflected classification 2: populated classification 3: sampling classification 4: spatial classification</p>

20.	<p>Q: The summary and presentation of data in tabular form with several non-overlapping classes is referred as</p> <p>Options</p> <p>1: nominal distribution  2: ordinal distribution  3: chronological distribution  4: frequency distribution</p>
21.	<p>Q: The largest value is 60 and smallest value is 40 and the number of classes desired is 5 then the class interval is</p> <p>Options</p> <p>1: 20  2: 4  3: 25  4: 15</p>
22.	<p>Q: The diagram used to represent group and ungrouped data is classified as</p> <p>Options</p> <p>1: breadth diagram  2: width diagram  3: bar diagram  4: length diagram</p>
23.	<p>Q: Histogram, pie charts and frequency polygon are all types of</p> <p>Options</p> <p>1: one dimensional diagram  2: two dimensional diagram  3: cumulative diagram  4: dispersion diagram</p>
24.	<p>Q: Which of the following is not a type of univariate frequency distribution</p> <p>Options</p> <p>1: Individual observation  2: Discrete frequency distribution  3: Continuous frequency distribution  4: Random frequency distribution</p>
25.	<p>Q: The method of classification of data in terms of class intervals in which both the lower limit and the upper limit of any class (class interval) are included in that class (class interval) is known as _____ method of classification</p>

	<p>Options</p> <p>1: exclusive</p> <p>2: inclusive</p> <p>3: equal</p> <p>4: unequal</p>										
26.	<p>Q: What is the arithmetic mean of 2, 8, 10, 6, 14?</p> <p>Options</p> <p>1: 5</p> <p>2: 6</p> <p>3: 7</p> <p>4: 8</p>										
27.	<p>Q: From the following frequency distribution, find the median class:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th><i>Cost of living index</i></th> <th><i>No. of weeks</i></th> </tr> </thead> <tbody> <tr> <td>1400-1550</td> <td>8</td> </tr> <tr> <td>1550-1700</td> <td>15</td> </tr> <tr> <td>1700-1850</td> <td>21</td> </tr> <tr> <td>1850-2000</td> <td>8</td> </tr> </tbody> </table> <p>Options</p> <p>1: 1400–1550</p> <p>2: 1550–1700</p> <p>3: 1700–1850</p> <p>4: 1850–2000</p>	<i>Cost of living index</i>	<i>No. of weeks</i>	1400-1550	8	1550-1700	15	1700-1850	21	1850-2000	8
<i>Cost of living index</i>	<i>No. of weeks</i>										
1400-1550	8										
1550-1700	15										
1700-1850	21										
1850-2000	8										
28.	<p>Q: We need _____ dimension(s) to table or plot a univariate (1-variable) frequency distribution</p> <p>Options</p> <p>1: one</p> <p>2: two</p> <p>3: three</p> <p>4: four</p>										
29.	<p>Q: Quota sampling is similar to _____ sampling.</p> <p>Options</p> <p>1: purposive</p> <p>2: convenience</p> <p>3: stratified</p> <p>4: cluster</p>										
30.	<p>Q: The number of possible samples of size 2 out of 5 population size in simple random sampling with replacement (SRSWR) is equal to</p> <p>Options</p>										

	<p>1: 10 2: 15 3: 20 4: 25</p>																														
31.	<p>Q: Which of the following method of sampling is not a part of 'restricted random sampling'?</p> <p>Options</p> <p>1: Lottery method 2: Stratified method 3: Systematic method 4: Cluster method</p>																														
32.	<p>Q: Consider simple random sampling without replacement (SRSWOR) from a population of size N. The number of samples of size n is</p> <p>Options</p> <p>1: <math>{}^N P_n</math> 2: <math>{}^N C_n</math> 3: <math>N^n</math> 4: N</p>																														
33.	<p>Q: Population census conducted by the government of India after every 10 years is an example of _____ data.</p> <p>Options</p> <p>1: Primary data 2: Secondary data 3: Structured data 4: Unstructured data</p>																														
34.	<p>Q: If <math>b_{yx} = 0.5</math> and <math>b_{xy} = 0.46</math>, then the value of coefficient of correlation (r) is</p> <p>Options</p> <p>1: 0.39 2: 0.48 3: 0.23 4: 0.25</p>																														
35.	<p>Q: From the given table, the MAE value is</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Week</th> <th>Demand</th> <th>Forecast</th> <th>Error</th> <th> Error </th> <th>Error<sup>2</sup></th> </tr> </thead> <tbody> <tr> <td>1</td> <td>100</td> <td>0</td> <td>-100</td> <td>100</td> <td>10 000</td> </tr> <tr> <td>2</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td>3</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td>Total</td> <td>100</td> <td>0</td> <td>-100</td> <td>100</td> <td>10 000</td> </tr> </tbody> </table>	Week	Demand	Forecast	Error	Error	Error <sup>2</sup>	1	100	0	-100	100	10 000	2	0	0	0	0	0	3	0	0	0	0	0	Total	100	0	-100	100	10 000
Week	Demand	Forecast	Error	Error	Error <sup>2</sup>																										
1	100	0	-100	100	10 000																										
2	0	0	0	0	0																										
3	0	0	0	0	0																										
Total	100	0	-100	100	10 000																										

Options 1: 100 2: 333.33 3: 133.33 4: 33.33
---

### Descriptive Questions

In a simple study about coffee habits in two towns A and B the following information is given

Town A: Females were 40%, total coffee drinkers were 45% and female non coffee drinkers were 20%.

Town B: males were 55%, male non coffee drinkers were 30% and female coffee drinkers were 15%

Present the data into a table format

What is diagrammatic representation of data explain advantages of it.

The data with regard to the output of gram and cost of seed and labour per hectare at ten farmers' fields, are as given below

Sr No	Cost of produce (Y) (RS/ha)	Cost of Seed (X1) (RS/ha)	Cost of Labour (Y) (RS/ha)
1	1127	235	128
2	840	236	82
3	735	238	205
4	570	241	71
5	462	238	110
6	614	233	130
7	916	235	200
8	460	190	170
9	1540	235	180
10	1065	243	165

- i) Fit the regression equation  $\hat{Y} = b_0 + b_1X_1 + b_2X_2$
- ii) Estimate the cost of produce per hectare given that  $X_1 = 230$  and  $X_2 = 125$
- iii) Test the significance of partial regression coefficients
- iv) Find the partial correlation coefficient  $r_{YX_2X_1}$

What do you mean by a questionnaire? What is the difference between a questionnaire and a schedule? State the essential points to be remembered in drafting a questionnaire.

What are the different methods of collection of data? Why are personal interviews usually preferred to questionnaire? Under what conditions may a questionnaire prove as satisfactory as a personal interview?

Perform simple linear regression

X	1	2	3	3	4	5
Y	8	4	5	2	2	0

Determine slope and intercept

Exercise on find correlation of the following data

x	9	8	7	6	5	4	3	2	1
y	15	16	14	13	11	12	10	8	9

Relationship of  $R^2$  and Adjusted  $R^2$  with additional of one independent variable

Using t-test check significance of independent variable.

What is the effect of  $R^2$  and Adjusted  $R^2$  for addition of new variable in multiple linear regression.

Obtain Partial correlation coefficients for following data

**Example 2:** From the following data, obtain  $r_{123}$ ,  $r_{132}$  and  $r_{231}$ .

$X_1$	20	15	25	26	28	40	38
$X_2$	12	13	16	15	23	15	28
$X_3$	13	15	12	16	14	18	14

1. Explain in details primary and secondary data.

How will you decide about the relative importance of various independent variables?

What is non-probability sampling and explain types of non-probability samplings.

From 10 observations on Price (x) and Supply (y) of a commodity, the following summary of figures were obtained.

$$\sum x = 130, \sum y = 220, \sum x^2 = 2288, \sum xy = 3467$$

Compute the line of regression of y on x and interpret the result. Estimate the supply when price is 16 units.

Let  $X_1, X_2, \dots, X_n$  be a random sample show that the sample mean

$$\hat{\theta} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \text{ is an unbiased estimator of } \theta = E(x_i)$$

Obtain Partial correlation coefficients for following data

**Example 2:** From the following data, obtain  $r_{123}$ ,  $r_{132}$  and  $r_{231}$ .

$X_1$	20	15	25	26	28	40	38
$X_2$	12	13	16	15	23	15	28
$X_3$	13	15	12	16	14	18	14

Frequency Distribution, Types of Univariate Frequency Distribution, Cumulative Frequency Distribution, Bivariate/Two-way classification of data, Cumulative frequency curve or ogive (i.e., more than ogive and less than ogive)

Methods to Check the Performance of Regression Models: MAE, MSE,  $R^2$ , MAPE (Moving Averages)

Sums on Point Estimate of the Population mean, Population Std Deviation, and Std. Error of the Estimate mean

Hypothesis Testing:

- a) Z test for Single Mean
- b) Z test for Difference of Mean

Explain 1.Test of significance 2.Level of significance 3.Simple hypothesis  
4.Composite Hypothesis

The manufacturer of a certain make of electric bulbs claims that his bulbs have a mean life of 25 months with standard deviation of 5 months. A random sample of 6 such bulbs gave the following values

Life of bulb in months 24,26,30,20,20,18

Is the manufacturer's claim valid at 1% level of significance?(Given that the table values of the appropriate test statistics at said level are 4.032,3.707 and 3.499 for 5, 6 and 7 degree of freedom respectively)

Explain in details MP and UMP-Test

Define MAPE, MAE, RMSE with formula and example