

Date: 04-03-2024

Time: 1 Hr.

Branch: CMPN

Semester: TE -VI

Subject: Quantitative Analysis

Marks: 30

Q. 1)	Attempt any Five (2 Marks Each)	CO	BL																																		
a)	What are the advantages of a tabular presentation of data	1	1																																		
b)	Explain Sturges Rule for the number of classes and size of the interval	1	3																																		
c)	Marks scored by 15 students are given as follows 6,8,10,12,13,15,19,21, 35, 28, 27, 33, 28,13, 22, 40 Convert the marks into a continuous series of a class interval of 10.	1	3																																		
d)	Explain Primary Data vs. secondary Data with an example(2 points)	2	3																																		
e)	Calculate Arithmetic and Geometric for a given dataset 1,3,9,27,82,241,720	1	3																																		
f)	Represent the following data by a Rectangle graph: <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <th>Items</th><th>Family</th></tr> <tr> <td>Food</td><td>1500</td></tr> <tr> <td>Misc</td><td>3050</td></tr> <tr> <td>Saving</td><td>450</td></tr> </table>	Items	Family	Food	1500	Misc	3050	Saving	450	2	2																										
Items	Family																																				
Food	1500																																				
Misc	3050																																				
Saving	450																																				
g)	Give an example of snowball sampling method	2	1																																		
h)	Advantages Non-Probability sampling methods	2	1																																		
Q. 2)	Attempt any two (5 Marks Each)																																				
a)	Calculate the mean and median of the grouped data <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <th>Score</th><th>40-49</th><th>50-59</th><th>60-69</th><th>70-79</th><th>80-89</th><th>90-99</th></tr> <tr> <td>Frequency</td><td>6</td><td>8</td><td>12</td><td>14</td><td>7</td><td>3</td></tr> </table>	Score	40-49	50-59	60-69	70-79	80-89	90-99	Frequency	6	8	12	14	7	3	1	1																				
Score	40-49	50-59	60-69	70-79	80-89	90-99																															
Frequency	6	8	12	14	7	3																															
b)	Construct a bivariate frequency distribution table of the marks obtained by students in English (X) and statistics (Y). <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <th>Marks in statistics(X)</th><td>37</td><td>20</td><td>46</td><td>28</td><td>35</td><td>26</td><td>41</td><td>48</td><td>32</td><td>23</td><td>20</td><td>39</td><td>47</td><td>33</td><td>27</td><td>26</td></tr> <tr> <th>Marks in English(Y)</th><td>30</td><td>32</td><td>41</td><td>33</td><td>29</td><td>43</td><td>30</td><td>21</td><td>44</td><td>38</td><td>47</td><td>24</td><td>32</td><td>21</td><td>20</td><td>21</td></tr> </table>	Marks in statistics(X)	37	20	46	28	35	26	41	48	32	23	20	39	47	33	27	26	Marks in English(Y)	30	32	41	33	29	43	30	21	44	38	47	24	32	21	20	21	1	3
Marks in statistics(X)	37	20	46	28	35	26	41	48	32	23	20	39	47	33	27	26																					
Marks in English(Y)	30	32	41	33	29	43	30	21	44	38	47	24	32	21	20	21																					
c)	For a given data set calculate less than more than and relative frequency. <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <th>Marks</th><th>No of students</th></tr> <tr><td>0-10</td><td>12</td></tr> <tr><td>10-20</td><td>6</td></tr> <tr><td>20-30</td><td>9</td></tr> <tr><td>30-40</td><td>16</td></tr> <tr><td>40-50</td><td>20</td></tr> <tr><td>50-60</td><td>15</td></tr> <tr><td>60-70</td><td>8</td></tr> <tr><td>70-80</td><td>8</td></tr> <tr><td>80-90</td><td>8</td></tr> </table>	Marks	No of students	0-10	12	10-20	6	20-30	9	30-40	16	40-50	20	50-60	15	60-70	8	70-80	8	80-90	8	1	3														
Marks	No of students																																				
0-10	12																																				
10-20	6																																				
20-30	9																																				
30-40	16																																				
40-50	20																																				
50-60	15																																				
60-70	8																																				
70-80	8																																				
80-90	8																																				
Q 3)	Attempt any one (10 Marks Each)																																				
a)	Explain different steps involved in data pre-processing Numerical data with regard to industrial diseases and deaths therefrom in Great Britain during the years 1935-39 and 1940-44 are given in a descriptive form: "During the quinquennium 1935-39, there were in Great Britain 1, 775 cases of industrial diseases made up of 677 cases of lead poisoning, 111 of other poisoning, 144 of anthrax, and 843 of gassing. The number of deaths reported was 20 p.c. of the cases for all the four diseases taken together, that for lead poisoning was 135, for other poisoning 25 and that for anthrax was 30. During the next quinquennium, 1940-44, the total number of cases reported was 2, 807. But lead poisoning cases reported fell by 351 and anthrax cases by 35. Other poisoning cases increased by 784 between the two periods. The number of deaths reported decreased by 45 for lead poisoning, but decreased only by 2 for anthrax from the pre-war to the post-war quinquennium. In the later period, 52 deaths were reported for poisoning other than lead poisoning. The total number of deaths reported in 1940-44 including those from gassing was 64 greater than in 1935-39" Represent the data in the tabular form	1	1																																		
b)	Explain different Probability and non-probability sampling methods with appropriate examples	2	1																																		
CO1	To understand basic statistical foundations for roles of Data Scientist																																				
CO2	To develop problem-solving skills																																				

Vidyalankar Institute of Technology
Semester VI – CMPN – Mid Semester Assessment – 2

Date: 01/04/2024

Quantitative Analysis

30 Marks/ 1 hour

1 Solve any five (2 marks each)

- | | | |
|---|--|-----|
| A | What is regression? Why it is called regression? | CO |
| B | What is an example of multiple regression in real life? | CO3 |
| C | A LR analysis produces the equation $Y = -3.2X + 7$. What does it indicate? | CO3 |
| D | How is regression analysis used in practical applications? | CO3 |
| E | What is the difference between simple linear regression and multiple linear regression? | CO3 |
| F | Predict sales when advertising expenditure (X1) is \$15,000 and the competitor's price (X2) is \$ 50. Let's assume the coefficients of the model are as follows: $\beta_0=50$ $\beta_1=4$ $\beta_2=-2$. | CO3 |
| G | How is the best-fitting line determined in simple linear regression? | CO3 |
| H | What is the "R-value" in regression? | CO3 |

2 Solve any one (10 marks each)

- | | | |
|---|---|-----|
| A | Find the two lines of regression from the following data.
a) Estimate the age of the husband if the age of the wife is 19.
b) Estimate the age of the wife when the age of the husband is 30. | CO3 |
|---|---|-----|

Age of Husband	Age of Wife
25	18
22	15
28	20
26	17
35	22
20	14
22	16
40	21
20	15

- | | | |
|---|--|-----|
| B | Find multiple linear regression equations of Y on X1 and X2. | CO3 |
|---|--|-----|

Y	X1	X2
4	15	30
6	12	24
7	8	20
9	6	14
13	4	10
15	3	4

3 Solve any one (10 marks each)

- | | | |
|---|--|-----|
| A | Find the regression equation to predict the glucose level given the age. | CO3 |
|---|--|-----|

Age	Glucose level
43	99
21	65
25	79
42	75
59	81
55	?

Vidyalankar Institute of Technology
Semester VI – CMPN- Mid Semester Assessment –3

Date: 22/04/2024

Quantitative Analysis

30 Marks/1 hour

- 1 Solve any two (5 marks each)
 - A What are the three characteristics of a good estimator? CO
 - B A certain type of storage battery lasts on average 3.0 years, with a standard deviation of 0.5 years. Assuming that the battery lives are normally distributed, find the probability that a given battery will last less than 2.3 years. CO4
 - C What is the critical region? Explain with diagram one-tailed test and two-tailed test. CO5
- 2 Solve any one (10 marks each)
 - A The sales data of an item in six shops before and after a special promotional campaign are as follows: CO5

Shops	A	B	C	D	E	F
Before campaign	53	28	31	48	50	42
After campaign	58	29	30	55	56	45

Can the campaign be judged to be a success? Test at a 5% level of significance.
 - B The Mean light of a sample of 10 electric bulbs is 1456 hours with a standard deviation of 423 hours. The second sample of 17 electric bulbs chosen from a different batch has a mean of 1280 hours with a standard deviation of 398 hours. Is there a significant difference between the mean of the two samples at a 5% level of significance? CO5
- 3 Solve any one (10 marks each)
 - A Suppose we have a dataset that shows the number of hours studied by six different students along with their final exam scores. CO4

Hours Studied	Exam Score
1	68
2	77
2	81
3	82
4	88
5	90

The prediction model is Exam Score = 66.615 + 5.0769*(Hours Studied). Calculate SST, SSR, SSE, and R-squared.
 - B Ten individuals are chosen at random, from a normal population and their weights (in kg) are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, and 71. Test the claim that the mean weight in the population is 66 kg at a 5% level of significance. CO4

Quantitative Analysis

Question Bank

MODULE 5: STATISTICAL INFERENCE

1.	Sums on SST, SSR, SSE, and R-squared
2.	Explain Point Estimation
3.	Explain Interval Estimate
4.	Explain the following Properties of Point Estimators 1. Bias 2. Consistency 3. Unbiasedness 4. Efficiency
5.	Explain Moments (method of point estimation)
6.	Explain maximum likelihood (method of point estimation)
7.	A random sample of $n=6$ has the elements 6, 10, 13, 14, 18, 20. Compute a point estimate of 1. Population mean 2. Population standard deviation

MODULE 6: TESTS OF HYPOTHESES

1.	Explain the concept of hypothesis. What is null hypothesis and alternate hypothesis?
2.	Explain in brief: Type I and Type II errors in hypothesis.
3.	Give the difference between a one tailed test and two tailed test in hypothesis.
5.	A medicine was found to be effective for 9 patients in 8 days on an average with standard deviation of 2.2 days. Another medicine administered to another group of 8 patients was found to be effective in 6 days on an average with standard deviation of 2.6 days. Use 5% level of significance to test the null hypothesis that the two medicines are equally effective.
6.	In a survey of buying habits, 400 women shoppers are chosen at random in supermarket 'A'. Their average weekly food expenditure is Rs. 250 with standard deviation of Rs. 40. For 400 women shoppers chosen at random in supermarket 'B', the average weekly food expenditure is Rs. 220 with standard deviation of Rs. 55. Test at 1% level of significance whether the

	average weekly food expenditure of the two populations of shoppers are equal.
7.	<p>A manufacturer claims that the average life of their electric light bulbs is 2000 hours. A random sample of 64 bulbs is tested and the life, x, in hours recorded. The results obtained are as follows:</p> $\Sigma x = 127\ 808 \quad \Sigma(\bar{x} - x)^2 = 9694.6$ <p>Is there sufficient evidence, at the 1% level, that the manufacturer is over estimating the length of the life of the light bulbs?</p>
8.	To verify whether a course in mathematics improves performance or not, two tests were given to 12 participants, one before and one after the course. Marks of the test paper before the course were 44, 40, 61, 52, 32, 44, 70, 41, 67, 72, 53, 72. Marks of the test after the course were 53, 38, 69, 57, 46, 39, 73, 48, 73, 74, 60, 70. Determine whether the course was useful or not.
9.	The Mean light of a sample of 10 electric bulbs is 1456 hours with a standard deviation of 423 hours. The second sample of 17 electric bulbs chosen from a different batch has a mean of 1280 hours with a standard deviation of 398 hours. Is there a significant difference between the mean of two samples?
10.	For a random sample of 10 persons fed on diet A, the increase in the weight for a certain period was 10, 6, 16, 17, 13, 12, 8, 14, 15, 9 kgs. For another sample of 12 persons fed on diet B, the increase in the weights for the same period was 7, 13, 22, 17, 15, 12, 14, 18, 8, 21, 23, 10 kgs. Test whether the two diets differ significantly as regards to increase in weights.
11.	A random sample is taken and the sample size is 25 The sample is normally distributed, then sample mean is 89 , and the standard deviation is 5.5. Find a 90% confidence interval for the population mean.
12.	What is the 95% confidence interval for population mean of a sample of 81 observation taken from a normal population with standard deviation of 5, if the sample mean is 40?
13.	Ten individuals are chosen at random from a population and their heights are found in inches. The heights are 63, 63, 64, 65, 66, 69, 69, 70, 70, 71. Discuss the suggestion that the mean and height of population is 66. Given for 9 degrees of freedom at 5% level of significance the value of t is 2.262.
14.	A machine is designed to pack edible oil in tins of 5 kgs. A random sample of 10 tins gave the average weight of a tin as 4.8 kg and standard deviation of 2 kg. Is the machine working properly? Given for 9 degrees of freedom at 5% level of significance the value of t is 2.262.
15.	<p>Cardiac patients were implanted pacemakers to control heartbeats. A plastic connector module mounts on top of pacemakers. Assuming standard deviation of 0.0015 inches and normal distribution, find 95% confidence level for mean size of the connector module.</p> <p>A random sample of 75 modules has an average of 0.31 inches. How large a sample is needed if we wish to be 95% confident that the sample mean will be within 0.0005 inches from the true mean?</p>

16.	Compare & Explain Null and Alternative Hypothesis																																										
17	<p>A breeder claims that his variety of cotton contains, at the most, 40 per cent lint in seed cotton. Eighteen samples of 100 grams each were taken, and after ginning the following quantity of lint was found in each sample.</p> <table><tr><td>Sample No. :</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr><tr><td>Quantity of Lint in 100 g sample:</td><td>36.3</td><td>37.0</td><td>36.6</td><td>37.5</td><td>37.5</td><td>37.9</td><td>37.8</td><td>36.9</td><td>36.7</td></tr><tr><td></td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td></tr><tr><td></td><td>38.5</td><td>37.9</td><td>38.8</td><td>37.5</td><td>37.1</td><td>37.0</td><td>36.3</td><td>36.7</td><td>35.7</td></tr></table> <p>Perform ‘ t’ Test for the hypothesis</p> <p>Note: Value of t tabulated = 2.567 at $\alpha = 0.01$</p>	Sample No. :	1	2	3	4	5	6	7	8	9	Quantity of Lint in 100 g sample:	36.3	37.0	36.6	37.5	37.5	37.9	37.8	36.9	36.7		10	11	12	13	14	15	16	17	18		38.5	37.9	38.8	37.5	37.1	37.0	36.3	36.7	35.7		
Sample No. :	1	2	3	4	5	6	7	8	9																																		
Quantity of Lint in 100 g sample:	36.3	37.0	36.6	37.5	37.5	37.9	37.8	36.9	36.7																																		
	10	11	12	13	14	15	16	17	18																																		
	38.5	37.9	38.8	37.5	37.1	37.0	36.3	36.7	35.7																																		
18.	<p>The table below gives the total income in thousand rupees per year of 36 randomly selected persons from a particular class of people.</p> <table><tr><th colspan="6">Income (thousand Rs.)</th></tr><tr><td>6.5</td><td>10.5</td><td>12.7</td><td>13.8</td><td>13.2</td><td>11.4</td></tr><tr><td>5.5</td><td>8.0</td><td>9.6</td><td>9.1</td><td>9.0</td><td>8.5</td></tr><tr><td>4.8</td><td>7.3</td><td>8.4</td><td>8.7</td><td>7.3</td><td>7.4</td></tr><tr><td>5.6</td><td>6.8</td><td>6.9</td><td>6.8</td><td>6.1</td><td>6.5</td></tr><tr><td>4.0</td><td>6.4</td><td>6.4</td><td>8.0</td><td>6.6</td><td>6.2</td></tr><tr><td>4.7</td><td>7.4</td><td>8.0</td><td>8.3</td><td>7.6</td><td>6.7</td></tr></table> <p>On the basis of the sample data, can it be concluded that the mean income of a person in this class of people is Rs. 10,000 per year? Note: Value of z tabulated = 1.96 at $\alpha = 0.05$</p>	Income (thousand Rs.)						6.5	10.5	12.7	13.8	13.2	11.4	5.5	8.0	9.6	9.1	9.0	8.5	4.8	7.3	8.4	8.7	7.3	7.4	5.6	6.8	6.9	6.8	6.1	6.5	4.0	6.4	6.4	8.0	6.6	6.2	4.7	7.4	8.0	8.3	7.6	6.7
Income (thousand Rs.)																																											
6.5	10.5	12.7	13.8	13.2	11.4																																						
5.5	8.0	9.6	9.1	9.0	8.5																																						
4.8	7.3	8.4	8.7	7.3	7.4																																						
5.6	6.8	6.9	6.8	6.1	6.5																																						
4.0	6.4	6.4	8.0	6.6	6.2																																						
4.7	7.4	8.0	8.3	7.6	6.7																																						
19.	Discuss your understanding about Null and Alternative Hypothesis																																										
20.	Write a short note on Errors in Hypothesis testing.																																										
21	What is critical region? Explain																																										
22.	Explain steps in solving testing of hypothesis problem																																										
23.	What is level of significance																																										
24.	<p>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</p> <p>Life of bulb in months 24,26,30,20,20,18</p> <p>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)</p>																																										
25.	Explain Type 1 and type 2 Errors in detail.																																										
26.	Explain One tailed and Two tailed Tests.																																										
27.	Explain Parametric and Non Parametric tests of hypothesis																																										
28.	Introduction to Hypothesis Testing																																										

29.	Types of Errors
-----	-----------------