Question Bank

Subject: Statistics for Artificial Intelligence and Data Science

Module1

1. From the following frequency distribution compute the standard deviation of 100 students

Mass in Kg.	No. of Students
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

2. Find Q1, Q2 and Q3 for the following data set and draw a box and whisker Plot. {2,6,7,8,8,11,12,13,14,15,22,23}.

3. Represent the adjoining distribution of marks of 100 students in the examination by a histogram.

Marks	Obtained	Number of Students
Less than	10	4
Less than	20	6
Less than	30	24
Less than	40	46
Less than	50	67
Less than	60	86
Less than	70	96
Less than	80	99
Less than	90	100

4. Differentiate between categorical data and numerical data.

5. Calculate the main deviation from the mean of the following distribution.

Marks	Number of students
0-10	5
20-20	8
20-30	15
30-40	16
40-50	6

6. The first of the two samples has 100 items with mean 15 and standard deviation 3. If the whole group has 250 items with mean 15.6 and standard deviation 13.44. Find the standard deviation of the second group.

Module 2

1. Define sampling. Explain in detail the different types of sampling.
2. Give the difference between sample mean and population mean.
3. State central limit theorem
4. Explain in detail the normal distribution.
5. State importance of normal distribution
6. Explain standard error with example
7. Explain bootstrap algorithm.
8. The time taken by the workers in performing a job by Method 1 and method to is given below.

Method I	20	16	26	27	23	22	
Method II	27	33	42	35	32	34	38

Do the data show that the variance of time distribution from population from which these examples are drawn do not differ significantly.

9. In a survey of 200 boys of which 75 wear intelligent, 40 had skilled fathers, while 85 of the unintelligible boys had unskilled fathers. Do they speakers support the hypothesis that's killed fathers having intelligent boys?(Chi square value = 3.841)

10. The mean lifetime of 100 light bulbs produced by a company with computer to be 1570 hours with the standard deviation of 120 hours. If you is the mean lifetime of all the bulbs produced by the company, test hypothesis $\mu = 1600$ hours against the alternative $\mu \mathrel{!=} 1600$ hours using a 5% level of significance.(z-test)

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

Module 3

1. Give comparison of null and alternate hypothesis.
2. Explain test of significance.
3. Explain level of significance.
4. Explain left till test and right tail test.
5. What is the difference between sample and population.
6. Mice with an average lifespan of 32 months will live up to 40 months when fed by a certain nutritious food. If 64 miles fed on this diet have an average lifespan of 38 months and standard deviation of 5.8 months, is there any reason to believe that lifespan is less than 40 months. (Z test).
7. Explain in detail the Chi-square test.
8. Explain ANOVA.
9. Explain in detail Type 1 error and type 2 error

Module 4

1. Explain in detail the cumulative distribution function.
2. Given below is the stem and leaf plot of the people who attended an election campaign from various places. Hari is trying to read the plot, but he has many doubts. Can you clarify them?
   (i) what are the data values of stem 2?
   (ii) how many values are less than 16?

Stem	Leaf
1	4 4 6
2	3 8
3	2
4	0 1

3. Explain and detail about trimmed mean.
4. Consider the following data set:
   {92,19,101,58,1053, 91,26,78,10,13,-40,101,86,85,15,89,89,28,-5,41}.
   Find the 20 percentage of trimmed mean.

Module 5

1. The trucking company wishes to test the average life of each of the 4 brands of tyres. The company uses all the branch on randomly selected trucks. The record showing the lives(thousands of miles) of tyres are as given in the table. Test the hypothesis that the average life for each brand of tyres is the same. Assume $\alpha = 0.01$.

Brand 1	Brand 2	Brand 3	Brand 4
20	19	21	15
23	15	19	17

2. The manufacturing company has purchase three new machines of different makes and wishes to determine whether one of them is faster than the others introducing a certain output. Five hardly production figures are observed at random from each machine and the results are given in the table. Use analysis of variance technique and determine whether the Machines or significantly different in their means speech.Use $\alpha = 0.05$.

Observations	Machine A1	Machine A2	Machine A3
	25	31	24
	30	39	30
	36	38	28
	38	42	25
	31	35	28

3. Write short note on survival analysis..
4. 7 random people were given 3 different drugs and for each person, the reaction time corresponding to the drugs were noted. Test the claim at the 5% significance level that all the 3 drugs have the same probability distribution.

	Drug A	Drug B	Drug C
1	1.24	1.50	1.62
2	1.71	1.85	2.05
3	1.37	2.12	1.68
4	2.53	1.87	2.62

5	1.23	1.34	1.51
6	1.94	2.33	2.86
7	1.72	1.43	2.86

5. With the following data on ml of potassium/quart in brands of drink, determine if there is a significant difference in the potassium content between brands.

Brand A	Brand B	Brand C
4.7	5.3	6.3
3.2	6.4	8.2
5.1	7.3	6.2
5.2	6.8	7.1
5.0	7.2	6.6

Use Kruskal Wallis Test.

6. Explain the statistics of One way ANOVA.
7. Explain in detail about Two way ANOVA.
8. Describe the Kruskal Wallis test algorithm

Module 6

1. Distinguish between regression and correlation.
2. What are the uses of regression analysis.
3. Discuss the properties of correlation coefficient
4. Explain in detail about coefficient of regression
5. Explain in detail about linear least square.
6. Distinguish between simple linear regression and multiple regression.