

Exam: Mid Term Odd Sem. 2020-21  
Course Name: Introduction to Data Science  
Course Code: CSE327  
Course Instructor: Shakti Balan  
Exam Date & Time : 28.09.20 / 10.00 am – 11.30 am



LNMIIT/B. Tech./CSE/Core/2020-21/ODD/CSE327/MT

**The LNM Institute of Information Technology**  
**Department of Computer Science & Engineering**  
**CSE 327 Introduction to Data Science**  
**Exam Type: Mid Term**

**Time: 90 min**

**28/09/2020**

**Max. Marks: 30**

NOTE: No doubt clarifications in the exam hall. If assumptions are to be made, make your own assumptions, state it and use it. If the assumptions are relevant and it makes sense it will be considered. **Bring your own calculators. Calculators should not be shared in the examination hall. All the best!**

1. Average amount of time required for the students to finish the IDS exam is 52 min with a SD of 10 min. What proportion of students would have finished the exam after 1 hour and 20 min.
2. A simple random sample of 105 undergraduates is taken at LNMIIT. The average Math 1 (M1) score of the sampled students is 54 with an SD of 12. Construct an approximate 95% confidence interval for the average M1 score of undergraduates at the university.
3. In order to test whether or not a random number generator is producing the digit "0" in the correct proportion (1/10), the generator will be run 6,500 times. You can assume that the runs are mutually independent and that each has the same probability  $p$  of producing "0". Construct a test that has a significance level of approximately 1%. [Note: "Construct a test" means "Come up with a decision rule." In the context of this problem, that means you have to say how you will use the number of 0's among your 6,500 results to decide between your hypotheses.]
4. For obese patients the Blood glucose levels have a mean of 100 with a standard deviation of 15. A medical researcher has an intuition that a diet high in raw cornstarch will have a positive or negative effect on blood glucose levels. A sample of 35 patients who have tried the raw cornstarch diet have a mean glucose level of 140. Test the hypothesis with 5% significance to check if the raw cornstarch had an effect or not.
5. The mean age for a sample of 1000 people from City A is 35 and the mean age for a sample of 600 people the City B is 28. It is learnt that the SD of age for City A is 5 and for City B is 7. Using these observations check if the mean of City A is significantly different from City B (Level of significance - 5%)?
6. In a group of 15 volunteers from different countries COVID19 vaccine test is performed. Three different vaccine were given, 5 volunteers were given Vaccine 1, 5 volunteers were given Vaccine 2 drug and 5 volunteers were given Vaccine 3. The level of antibodies in the blood is measured. The level of antibodies in the volunteers are as follows (in mg/dl)

S.No	Vaccine-1	Vaccine-2	Vaccine-3
1	710	396	215
2	752	524	312
3	800	752	411
4	991	359	233
5	695	861	725

Apply one-way ANOVA to test the hypothesis that the effectiveness of all three vaccine have different effects on the volunteers. The level of significance is  $\alpha = 0.05$ . The critical value of F score is 3.885294

All the best!

7. A shop S is selling the following items: Apple, Banana, Bread, Milk, Biscuits, and Sugar. There is a customer A who has bought Apple, Banana, Milk and Sugar, a customer B who has bought Apple, Banana, Bread and Sugar and a customer C has bought Banana, Milk, Biscuits and Sugar. S wants to find which customers B or C is closest to A. Which similarity measure you will use to answer this question. Why? Use the same to find whether B is closer to A or C is closer to A.

For the same problem given above, if we replace the given data by customer A bought 5 pieces of Apple, 2 pieces of Banana, 1 Liter Milk and 2 Kg Sugar, a customer B who has bought 3 Apples, 5 Bananas, 4 Breads and 2 Sugar and a customer C has bought 5 Banana, 1 Milk, 1 Biscuit and 2 Sugar. If S wants to find which customers B or C is closest to A, what similarity measure you will use and why? No need to calculate the similarity measure.