Reg. No.: .....

Name: .....

University of Kerala

W7096

Third Semester FYUGP Degree Examination, November 2025

Discipline Specific Core Course

STATISTICS

UK3DSCSTA210 - DATA ANALYSIS USING R- II

Academic Level: 200-299

2024 Admission

Time: 1 Hour 30 Minutes(90 Mins.) **Max. Marks: 42**

Part A. 6 Marks.Time:6 Minutes.(Cognitive Level:Remember(RE)/Understand(UN)) Objective Type. 1 Mark Each.Answer all questions

Qn No.	Question	CL	CO
1	A non-parametric alternative to ANOVA is … .	RE	4
2	Name a statistical test to compare the means of two dependent samples, whose population SD's are unknown.	RE	3
3	State whether the following statement is TRUE or FALSE: "A t-test can be applied only when the normality assumption is satisfied."	UN	3
4	The R command for one sample t test is	UN	3
5	What is the null hypothesis for a goodness of fit test?	UN	4
6	Which test is used for testing the equality of variances of two populations?	UN	3

Part B.8 Marks.Time:24 Minutes.(Cognitive Level:Understand(UN)/Apply(AP))Short Answer. 2 marks each.Answer all questions

Qn No.	Question	CL	CO
7	Write the use of R function shapiro.test()	UN	2
8	What do you mean by non-parametric tests?	UN	4
9	A 95% confidence interval for the average number of hours students study per week is given as (12.8, 17.2). Determine the sample mean and margin of error.	AP	2
10	A hospital is conducting a study to compare the effectiveness of two pain medications, Drug A and Drug B, in reducing post-operative recovery time (in days) for patients who underwent the same type of surgery. Doctors want to determine if there is a significant difference in average recovery time between the two groups of patients using different medications. Formulate the null and alternative hypotheses.	AP	3

Qn No.	Question	CL	CO

Part C. 28 Marks.Time:60 Minutes (Cognitive Level:Apply(AP)/Analyse(AN)/Evaluate(EV)/Create(CR)) Long Answer.7 marks each.Answer all 4 Questions choosing among options * within each question

Qn No.	Question	CL	CO
11	A) Write a program in R to compute a 95% CI for mean of a normal population with SD 4 from the following data: 150, 155, 160, 148, 152, 157, 149, 156, 159, 151, 148, 152. Justify your choice of the Statistical method. Also interpret the following output for the CI: [1] 150.8202 155.3465 attr(,"conf.level") [1] 0.95 OR B) A government survey of 800 households found that 240 of them are living below the poverty line. i) Compute the sample proportion. ii) Write a program in R to estimate the 95% confidence interval for the true proportion of households living below the poverty line in the population. iii) Also interpret the following output. [1] 0.2686591 0.3332900 attr(,"conf.level") [1] 0.95	AP	2, 2
12	A) Imagine a candy company claims that its bags of mixed candies contain the following proportions of different colours: Red: 20%, Green: 30%, Blue: 40% and Yellow: 10%. We want to test this claim and collected a random sample of 100 candies from a bag and count the number of each colour: Red: 15, Green: 35, Blue: 45 and Yellow: 5. Examine the scenario and answer the following. i) Formulate the hypotheses ii) Write a suitable R program to perform the test. iii) Some major results in the output of the desired R program are given below. Comment on the results. p-value = 0.2133 OR B)	AN	4, 3

Qn No.	Question	CL	CO
	Given the test scores of 10 students before and after giving a particular training are as follows. **Pre - score** 77 56 64 60 57 53 72 62 65 66 **Post -score** 88 74 83 68 58 50 67 64 74 60 Write the R code to test whether $H_0 : \mu1 = \mu2$ against $H_1 : \mu1 \neq \mu2$		
13	A) A large hospital system is comparing the effectiveness of two different medications for lowering cholesterol. They want to determine if there is a significant difference in the mean reduction of cholesterol levels between Medication A and Medication B. Based on extensive past research, the population standard deviation for cholesterol reduction with Medication A is known to be 25 mg/dL and for Medication B is 30 mg/dL. The hospital conducts a study, randomly selecting 12 patients to receive Medication A and 13 patients to receive Medication B. After a specified period, they measure the cholesterol reduction in each group and the results are given below. Medication A 82 125 116 151 96 138 102 154 121 115 106 105 Medication B 68 91 137 123 119 113 76 128 177 81 154 127 130 It is also known that both the populations are normal. The hospital wants to test if Medication A leads to a greater reduction in cholesterol than Medication B. Evaluate the scenario and answer the following. i) Formulate the hypotheses ii) Write a suitable R program to perform the test. iii) Some major results in the output of the desired R program are given below. Comment on the results. p-value = 0.9745 99 percent confidence interval: -28.01831 28.72344 sample estimates:	EV	3, 3

Qn No.	Question	CL	CO
	mean of x mean of y 117.5833 117.2308 OR B) John believes that 50% of first-time brides in the United States are younger than their grooms. She performs a hypothesis test to determine if the percentage is the same or different from 50%. He took a sample of 100 first-time brides and 53 reply that they are younger than their grooms. We want to perform the hypothesis test, at 1% level of significance. Evaluate the scenario and answer the following. i) Formulate the hypotheses ii) Write a suitable R program to perform the test. iii) Some major results in the output of the desired R program are given below. Comment on the results. p-value = 0.6171 99 percent confidence interval: 0.3988704 0.6572558		
14	A) i) Create a suitable example for conducting one-way ANOVA. The scenario must be different from those are being mentioned in the rest of the question paper. ii) Formulate the hypotheses. iii) Write a suitable program in R for the test. OR B) i) Create a suitable example for conducting Mann-Whitney-Wilcoxon test. The scenario must be different from those are being mentioned in the rest of the question paper. ii) Formulate the hypotheses. iii) Write a suitable program in R for the test.	CR	3, 4