CSC 215-2081(NEW) ✡

Tribhuvan University
Institute of Science and Technology
2081
✡

Bachelor Level / Second Year/ Third Semester/ Science
Computer Science and Information Technology (STA 215)
(Statistics II)
(NEW COURSE)

Full Marks: 60
Pass Marks: 24
Time: 3 hours

Candidates are required to give their answers in their own words as for as practicable. All notations have the usual meanings. The figures in the margin indicate full marks.

Group A

Attempt any TWO questions. **(2×10=20)**

1. Explain the purpose of applying multiple regression analysis. Following table shows the scores (Y) made by ten assemblies –line employees on a test design to measure the job satisfaction. It also shows the scores made on an aptitude test (X_1) and number of days absent (X_2) during the past year (excluding vacation).

Y	64	70	85	50	60	72	75	55	80	70
X_1	6	6	9	5	6	7	8	6	8	6
X_2	2	1	0	8	2	1	5	9	1	1

i) Find the multiple regression equation for the sample data.
ii) Interpret the value of regression coefficients b_1 and b_2.
iii) Estimate the score of an employee whose aptitude test score is 9 and number of absent days are 6.

2. A management consulting company presents a 3-day seminar on project management to various clients. The seminar is basically the same each time it is given. However, sometimes it is presented to high-level managers, sometimes to mid-level managers, and sometimes to low-level managers. The seminar facilitators believe evaluations of the seminar may vary with the audience. Suppose the following data are some randomly selected evaluation scores from different levels of managers after they have attended the seminar. The ratings are on a scale from 1 to 100, with 100 being the highest score. Use a one-way ANOVA to determine whether there is a significant difference in the evaluation according to manager level. The following table gives the scores to the various clients due to different manager levels.

High Level	Mid Level	Low Level
85	90	55
75	100	75
85	95	80
60	85	75
70	90	
	75	

3. Explain, stating clearly the assumptions involved in the t-test for testing the significance of difference between two sample means. Measurements of the left-and-right-hand gripping strengths of 8 left-handed writers are recorded:

Person	1	2	3	4	5	6	7	8
Left hand	112	131	142	90	125	130	95	90
Right hand	104	136	135	86	132	120	86	85

Do the data provide strong evidence that the people who write with left hand have greater gripping strength in the left hand than they do in the right hand? Use $\alpha = 0.05$.

Group B

Attempt any EIGHT questions. **(8×5=40)**

4. What do you understand by design of experiments? Give the layout of a Latin Square Design. Explain why the number of treatments tested in a Latin Square Design should not be less than 3?

5. Based on interviews of couples seeking divorce, a social worker compiles the following data related to the period of acquaintanceship before marriage and the duration of marriage:

Acquaintanceship before marriage	Duration of marriage		Total
	Less than or equals to 5 years	More than 5 years	
Below 0.5 year	15	7	22
0.5 – 1.5 years	26	22	48
Over 1.5 years	19	11	30
Total	60	40	100

Perform a test to determine if the data substantiate an association between the duration of a marriage and the acquaintanceship prior to marriage. Use 5% level of significance.

6. Define confidence level in estimation. A quality control inspector collected a random sample of 400 tubes of toothpaste from the production line and found that 20 of the tubes had leaks from the tail end. Construct 96% confidence interval for the percentage of all the toothpaste tubes that had leakage and interpret the result.

7. The mean drying time of a brand of spray paint is known to be 122 seconds. The research division of the company that produces this paint contemplates that adding a new chemical ingredient to the paint accelerate the drying process. To investigate this conjecture, the paint with the chemical additions is sprayed on 50 surfaces and the drying time is recorded. The mean and standard deviation of drying time computed from these recoded are found as 116 seconds and 16.8 seconds respectively. Do these data provide strong evidence that the mean drying time is reduced by the addition of the new chemical? Use 5% level of significance. Also find p-value.

8. Explain the concept of multiple and partial correlation coefficients. Consider three variables X_1, X_2 and X_3. If $r_{12} = 0.40$, $r_{23} = 0.50$ and $r_{13} = 0.6$ find $R_{1.23}$ and $r_{23.1}$.

9. Discuss the concept of level of significance in hypothesis testing. A manufacturer of laptop provides a particular model in one of three colors. Of the first 100 laptops sold, it is noted that 80 were the first color. Can you conclude that more than two third of all the customers have a preference for the first color? Use 5% level of significance.

10. A multiple regression equation yields the following results:

Source	Sum of square	Degree of freedom
Regression	740	2
Error	510	17

i) What is the total sample size?
ii) How many independent variables are being considered?
iii) Compute the coefficient of determination and interpret its value.
iv) Compute the standard error of estimate.
v) Test the hypothesis that the overall fit of the model is significant or not. Assume $\alpha = 0.05$.

11. Explain briefly the queuing theory. Customers arrive at a one-man barber shop according to Poisson process with mean inter arrival time of 12 minutes. Customer spends an average of 10 minutes in the barber's chair. What is the expected number of customers in the barber shop in the queue?

12. Write short note on:
a) Central limit theorem.
b) Determination of required sample size to estimate population proportion.