Tribhuvan University
Institute of Science and Technology
2080

Bachelor Level / third-semester / Science
Computer Science and Information Technology(STA215)
Statistics II

Full Marks: 60 + 20 + 20
Pass Marks: 24 + 8 + 8
Time: 3 Hours

Candidates are required to give their answers in their own words as far as practicable.
The figures in the margin indicate full marks.

Group A

Attempt any two questions.

A computer manager needs to know how efficiency of her new computer program depends on the size of incoming data and how many tables are used to arrange each data set. Efficiency will be measured the number of processed requests per hour. Applying the program to data set of different sizes and number of tables are used, she gets the following results.

Processed requests, Y	16	26	17	41	50	55	40
Data size, (gigabytes), X_1	15	10	10	8	7	7	6
Number of tables, X_2	1	2	10	10	20	20	4

The regression equation obtained is $Y = 52.7 - 2.87 X_1 + 0.85 X_2$.

Total sum of square = 1452

1 Sum of square due to regression = 1143.3.

a) Interpret the values of regression coefficients b_1 and b_2.

b) Test the significance of the regression model at 0.05 level of significance.

c) Is there significant relationship between processed request and number of tables at 0.05 level of significance? Given standard error of b_2=0.55.

d) What percentage of variation of processed requests is explained by data size and number of tables?

e) Compute standard error of estimate.

f) Estimate the number of processed requests if data size is 9 gigabytes and number of tables used are 8.

In an experiment to determine which of three different missile systems is preferable, the propellant burning rate is measuresd. The data after coding are given in the table. Use Kruskal-Wallis test (significance level of 0.01) to test the hypothesis that the propellant burning rates are same for three missile systems.

2

Missile system I	22.3	16.7	22.7	19.3	18.5		
Missile system II	23.4	19.5	17.5	20.8	16.0	19.9	
Missile system III	18.4	19.5	17.8	18.0	19.6	22.8	17.1

3 What is Latin Square Design? Under what conditions can this be used? Give lay out and analysis of Latin Square Design.

Group B

Attempt any eight questions

4 What do you understand by estimation? If we want to determine average mechanical aptitude of a large group of workers, how large a random sample is needed to be able to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 2.0 points? Assume that population standard deviation is 30.

5

A random sample of students is asked their opinion on proposed core curriculum change. The results are as follows:

Class	Options	
	Favoring	Opposing
Freshman	125	80
Sophomore	60	140
Junior	50	60
Senior	40	55

Test the hypothesis that opinion on the change is independent of class standing. Use 0.01 significance level.

6

Define Central limit theorem. The life of a certain brand of an electric bulb may be considered a random variable with mean 1350 hours and standard deviation 55o hours. Using central limit theorem, find the probability that the average life time of 100 bulbs exceeds 1440 hours.

7

Define multiple correlation. In a trivariate distribution X_1, X_2, and X3, the simple correlation coefficients are given as r_{12}= 0.5, r_{23}=0.6 and r_{13}=0.7. Find

i. partial correlation coefficient between X_1 an X_2 keeping X_3 constant.

ii. multiple correlation coefficient assuming X_1 as dependent variable.

8

What do you understand by Design of Experiment? Prepare one way analysis of variance table and carry out the test for the significance of difference in the average yields between different varieties of seed. Given:

Total sum of squares = 258

Sum of square between varieties of seed = 50

Total number of observations = 20

9

Define type I and type II error in testing of hypothesis. It is claimed that Samsung and Huwaei mobiles are equally popular in Kathmandu. A random sample of 600 people from Kathmandu showed 350 have Samsung mobile. Test the claim at 5% level of significance.

10

Customers of certain Internet service provider connect to the internet at the average rate of 10 new connections per minute. Connections are modelle by binomial counting process.

a. What frame length gives the probability 0.1 of an arrival during given frame?

b. Find the mean and variance for the number of seconds between two consecutive connections.

11

Write short notes on any two:

a. Difference betweeen parametric and non-parametric test.

b. Required assumptions for linear regression model.

c. Stochastic process.