

Tribhuvan University
Institute of Science and Technology

2079



Bachelor Level / Second Year/ Third Semester/ Science
Computer Science and Information Technology (STA 210)
(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 far 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. What are the required conditions for error variable in multiple regression analysis? The Internal Revenue Service (IRS) is trying to estimate the monthly amount of unpaid taxes discovered by its auditing division. The IRS estimated this figure on the basis of field auditing labor hours and numbers of hours of its computers are used. The table given below presents these data for the last ten months

Month	(X ₁) Field Audit Labor Hours in 100	(X ₂) Computer Hours in 100	(Y) Actual Unpaid Taxes Discovered million of dollars
Jan	45	16	29
Feb	42	14	24
Mar	44	15	27
April	45	13	25
May	43	13	26
June	46	14	28
Jul	44	16	30
Aug	45	16	28
Sept	44	15	28
Oct	43	15	27

Given: $\sum YX_1 = 12005$, $\sum YX_2 = 4013$, $\sum X_1X_2 = 6485$, $\sum Y^2 = 7428$, $\sum X_1^2 = 19461$, $\sum X_2^2 = 2173$.

- i. Develop the estimating equation best describing these data.
 - ii. Interpret the value of regression coefficients.
 - iii. Estimate the actual unpaid tax for field audit labor hours is 4200 and computer hours is 1600 hours.
2. ii) What do you understand by "Design of an Experiment"? Physicians depend the laboratory test results when managing the medical problems such as diabetes or epilepsy. In a uniformity test glucose tolerance, three different laboratories were each sent $n_i = 5$ identical blood samples from a person who had drunk 50 mg. of glucose dissolved in water. The laboratory results (mg./dl) are listed here:

Lab 1	Lab2	Lab3
12.1	9.3	10.0
11.7	11.1	10.5
10.9	10.7	10.1
10.2	10.9	11.0
10.6	9.0	10.4

Do data indicate a difference in the average readings for the three laboratories? Use $\alpha = .05$.

3. Define Type I and Type II error in testing of hypothesis. A psychologist wishes to verify that a certain drug increases the reaction time to given stimulus. The following reaction times (in tenth of seconds) were recorded before and after injection of the drug for each of four subjects:

Subject		1	2	3	4
Reaction Time	Before	7	2	12	12
	After	13	3	18	13

Test at the 5% level of significance to determine whether the drug significantly increases reaction time.

Group B**Attempt any Eight questions.****[8×5=40]**

4. The following ANOVA summary table was obtained from a multiple regression model with two independent variables.

Source of Variation	Sum of square	Degree of freedom	Mean sum of square	F -value
Regression	12.62	2	?	?
Error	0.78	12	?	
Total	13.40	14		

- Determine the mean sum of square due to regression, the mean sum of square due to error, and F - value.
- Test the significance of the overall regression model at 5% level of significance.
- Compute coefficient of determination and interpret its value.
- Find standard error of estimate.

5. What do you mean by non parametric test? Write down advantages of non parametric tests over the parametric tests.

6. Bank of Nepal recorded the sex of first 30 customers who appeared last Mon day with notation M M F M M F M F F M M F F F M F F M F F M M M F F . At the 0.05 level of significance, test the randomness of the sequence.

7. Social media users use a variety of devices to access social networking; mobiles phones are increasingly popular. However, is there a difference in the various age groups in the proportions of social media users who use their mobile phone to access social networking? A study showed the following results for the different age groups.

Use mobile phones to access social net working?	Age		
	18-34	35-64	65+
Yes	60	37	14
No	40	63	86

At the 0.05 level of significance, is there evidence of a difference among the age groups with respect to use of mobile phone for accessing social networking?

8. It is claimed that Samsung and Redmi mobiles are equally popular in Kathmandu. A random sample of 500 people from Kathmandu showed 300 have Samsung mobile. Test the claim at 5% level of significance.

9. An effort to estimate the mean amount per customer for dinner at a major Atlanta restaurant, data were collected for a sample of 49 customers and sample mean is found as \$ 24.80. Assume population standard deviation is \$5.
- Compute standard error of mean.
 - Find 95% confidence interval estimate for the population mean.

10. Define Markov chain and describe its characteristics.

11. What are the basic concepts of queuing theory? In a super market, the average arrivals rate of customer is 10 per every 30 minutes following Poisson process. The average time taken by the cashier to list and calculate the customers purchase is 2.5 minutes following exponential distribution. What is the probability that queue length exceeds 6. What is the expected time spent by customer in the system?

12. Write short notes on following.

- Partial and multiple correlation coefficient.
- Properties of good estimator