

TRIBHUVAN UNIVERSITY

Institute of Science and Technology

2065

Bachelor Level/ First Year/ First Semester/ Science

Computer Science and Information Technology (Stat. 108)

(Statistics I)

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.

Group A

Attempt any two:

(2x10=20)

1. Differentiate simple random sampling and stratified random sampling. Show that
 - (i) In simple random sampling without replacement (SRSWOR), the sample mean is an unbiased estimate of the population mean.
 - (ii) In SRSWOR, the variance of the sample mean is given by

$$Var(\bar{y}_n) = \frac{S^2}{n} \cdot \frac{N-n}{N}$$

2. Test the hypothesis of no difference between the ages of male and female employees of a certain company using the Mann-Whitney U test for the sample data. Use the 0.10 level of Significance.

Males	31	25	38	33	42	40	44	26
Females	44	30	34	47	35	32	35	47

3. Suppose you are given following information with $n = 28$.

Multiple regression model $\hat{Y} = 5 + 18X_1 + 20X_2$

Sample size (n) = 28

Total sum of squares (TSS) = 250

Sum of squares due to error (SSE) = 100

Standard error of regression coefficient of X_1 (Sb_1) = 3.2

Standard error of regression coefficient of X_2 (Sb_2) = 5.5

- i. Predict the value of Y for $X_1 = 15$ and $X_2 = 25$
- ii. Test the significance of regression coefficient of X_2
- iii. Compute the coefficient of multiple coefficient of determination.

Group B

Answer any eight questions:

(8x5=40)

4. Show that in two stage sampling with SRSWOR at both stages, \bar{y} is an unbiased estimator of \bar{Y} .
5. What do you mean by partial correlation coefficient? State the relationship between simple and partial correlation coefficients when there are three variables. If $r_{12} = 0.5$, $r_{23} = 0.1$ and $r_{13} = 0.4$, compute $r_{12.3}$ and $r_{23.1}$.
6. Define cluster sampling with sample mean and variance of sample mean.
7. For what conditions non parametric test is used. Explain some important non parametric test.

8. The weights of 4 people before they stopped smoking, in kilogram, are as follows:

Before	66	80	69	52	75
After	71	82	68	56	73

Use the signed-rank test for paired observations to test the hypothesis, at 0.05 level of significance, that giving up smoking has no effect on a person's weight against the alternative that one's weight increases if he or she quits smoking.

9. A random sample of 15 adults living in a small town is selected to estimate the proportion of voting favoring a certain candidate for mayor. Each individual was also asked if he or she was a college graduate. By letting Y and N designate the responses of "yes" and "no" to the education question, the following sequence was obtained:

N N N N Y Y N Y Y N Y N
N N N

Use the runs test at the 0.1 level of significance to determine if the sequence supports the contention that the sample was selected at random.

10. In an experiment to study the dependence of hypertension on smoking habits, the following data were taken on 180 individuals:

	Non smokers	Moderate smokers	Heavy smokers
Hypertension	21	36	30
No hypertension	48	26	19

Test the hypothesis that the presence or absence of hypertension is independent of smoking habits. Use a 0.05 level of significance.

11. Define dummy variable. What condition should be necessary for fitting logistic regression?

12. Suppose the residuals for a set of data collected over 9 consecutive time periods are as follows:

Time period:	1	2	3	4	5	6	7	8	9
Residuals:	-2	-3	+2	-1	0	1	4	-2	1

Compute the Durbin Watson statistics. At the 0.05 level of significance, is there evidence of autocorrelation among the residuals?

13. Describe a multiple regression model with its assumption. Also describe the method of obtaining its parameters.