

**TRIBHUVAN UNIVERSITY**  
Institute of Science and Technology  
2067

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)

- Describe in detail stratified random sampling method for drawing a random sample of size  $n$  from a population of size  $N$ . Write down the expression for an unbiased estimator  $\bar{y}_{st}$  of population mean  $\bar{Y}$  and derive an expression for  $Var(\bar{y}_{st})$  when samples were drawn from each stratum by adopting simple random sampling without replacement method. Also find the  $Var(\bar{y}_{st})$  under the scheme of proportional allocation of sample sizes to strata.
- Write down the rationale and method of Wilcoxon matched-pairs signed rank test. Seven prospective graduate students took a test twice with the following scores.

First attempt	470	530	610	440	600	590	580
Second attempt	510	550	600	490	585	620	598

Compute the value of  $T^+$  where  $T^+$  is the sum of ranks of the positive differences (second attempt – first attempt) using  $T^+$  as test statistic carry out the test of the following hypothesis at level 0.05.

$H_0$  : there is no statistical difference between the first and second attempt score

$H_1$  : second attempt score tends to be larger than the first attempt score.

- To study the effect of age ( $X_1$  in years) and weight ( $X_2$  in lbs) on systolic blood pressure ( $Y$  in mm Hg), the data were recorded for a sample of 15 adult males. The estimated regression model based on data is described below in the box where figures within parenthesis are standard error of the estimate. Further computation shows that

$$\sum (Y_i - \bar{Y})^2 = 1835.7 \text{ and } \sum (Y_i - \hat{Y}_i)^2 = 1101.3$$

$\hat{Y} =$	27.4+	0.22 $X_1$ +	0.56 $X_2$
	(24.68)	(0.248)	(0.155)

- Explain the meaning of the estimated slope regression coefficients of the model.

- (b) What value of Y would you predict if  $X_1 = 55$  and  $X_2 = 175$ ?
- (c) Compute the value of  $R^2$  and interpret it.
- (d) Carry out the overall goodness-of-fit test of the model at 5% level of significance.
- (e) Test the significance of slope regression coefficients at 5% level of significance.

**Group B**

**Answer any eight questions:**

(8x5=40)

4. Describe in detail systematic sampling method when  $N = k \times n$ . Describe problems that will arise in systematic sampling method when  $N \neq k \times n$ .
5. If  $V_{srswr}$  and  $V_{srswor}$  correspondingly denote that variance of unbiased estimator of the population mean under simple random sampling with and without replacement method, then show that  $(V_{srswr} - V_{srswor}) = \frac{n-1}{Nn} S^2$  and write your conclusion based on this result.
6. Consider the problem of determining if a die is fair or not. For this a die is rolled for 60 times and observed the following outcomes.

Side	1	2	3	4	5	6	Total
Number of times observed	8	9	13	7	15	8	60

Test the hypothesis  $H_0$  : the die is fair, that is, all sides have 1/6 chance of appearing against  $H_1$  : the die is unfair at level 0.05.

7. Describe the method of Mann Whitney test.
8. Suppose that an IQ test is given to eleven randomly selected pairs consisting of one brother and one sister from the same family. To test the null hypothesis that this sample was drawn from a population in which the median IQ of a brother and sister do not differ against the alternative hypothesis that the sister would score higher than brother. IQ scores are summarized below.

Sister's score	129	111	117	120	116	101	107	127	105	123	113
Brother's score	115	108	123	104	110	98	106	119	95	130	101

Using sign test carry out the above said hypothesis at 5% level of significance.

9. Describe rationale and method of Kriskal – Wallis one-way ANOVA test.
10. Suppose in a multiple regression model problem, the ANOVA table is as follows. How many independent variables are in the model? What is the sample size? What is the value of  $R^2$ ? Carry out the overall goodness-of-fit test of the model at 5% level of significance.

Source	SS	Df
Regression	36	2
Error	64	32

11. Explain the meaning of multicollinearity. How do you detect the problem of multicollinearity in multiple regressions?
12. Describe the Cobb-Douglas production function model with its application.
13. Define partial correlation coefficient. If  $r_{12} = 0.33$ ,  $r_{13} = 0.40$ , and  $r_{23} = 0.76$ , then compute  $r_{13.2}$  and  $r_{23.1}$ .

bsscsitblog.blogspot.com