Exam.Code:0041
Sub. Code: 1004

2022

B.Sc. (Hons.) Bio-Informatics

Third Semester

BIN-3004: Statistical Methods

Time allowed: 3 Hours

Max. Marks: 60

NOTE: *Attempt five questions in all, including Question No. 1 which is compulsory and selecting two questions from each Unit.*

x-x-x

Q1. Answer the following questions: **(2*6=12)**

a) Let b_{yx} and b_{xy} stand for the coefficients of regression of Y on X and X on Y respectively. Show that:

$$r_{xy} = \sqrt{b_{yx} \times b_{xy}}$$

b) Write the formula used to calculate the following: $\bar{x}, s^2, S^2, S.E.(\bar{x})$? Write their name also.

c) Examine the significance of r for the following situation when n = 8 and r = 0.2.

d) Derive the two normal equations used to determine the least squares line of best fit $Y_c = a + bX$.

e) State how the technique of grouping of experimental units into homogeneous blocks helps to control the error.

f) A population consists of the numbers 12, 14, 15 and 16. Enumerate all possible samples of size three which can be drawn from the population (i) with replacement and (ii) without replacement.

UNIT-I

Q2. (a) Answer the following: **(3,3)**

i. Prove that two independent variables are uncorrelated. By giving an example, show that the converse is not true. Explain the reason?

ii. What are the assumptions on which Karl Pearson's coefficient of correlation is based? Show that it is not affected by the change of origin and scale.

(b) Calculate the Karl Pearson's coefficient of correlation between X and Y from the following data and calculate probable error and interpret its value. Assume 69 and 112 as the mean value for X and Y respectively.

X	78	89	99	60	59	79	68	61
Y	125	137	156	112	107	136	123	108

Comment on the result. **(6)**

Q3. (a) Define regression. What are the assumptions of least square regression method? Distinguish between regression and correlation. **(6)**

(b) Obtain the equation for the line of regression of Y on X and X on Y. What are the properties of regression coefficients? What is the range of regression coefficients? **(6)**

Q4. (a) What do you mean by confidence interval? How will you determine the confidence interval for normal population mean? **(5)**

(b) A random sample of 800 units from a large consignment showed that 200 were damaged. Find 95% and 99% confidence limits for the population proportion of damaged units in the consignment. **(7)**

Q5. (a) What are the factors responsible for the determination of sample size for estimation of mean and proportion? How do you determine a sample size in both the cases? **(7)**

P.T.O.

(b) It is known that the population standard deviation in waiting time for L.P.G. gas cylinder in Delhi is 15 days. How large a sample should be chosen to be 95% confident, the waiting time is within 7 days of true average. **(5)**

UNIT-II

Q6. (a) Define the hypothesis and purpose of hypothesis testing. Write the complete steps of hypothesis testing. Write the null and alternative hypothesis of simple ANOVA and explain rejection of null hypothesis. **(6)**

(b) An ecologist collected organisms of a particular species from three different locations and counted the number of female organism in each sample. The results are given below in table

Location	**A**	**B**	**C**
No. of females	44	86	110
Total no. of samples	100	200	200

Test at 5% level of significance, whether the proportions of females differ significantly between the locations. **(6)**

Q7. (a) What do you mean by unpaired t-test? Write down the steps involved in unpaired t-test. **(6)**

(b) In hospital 480 female and 520 male babies were born in a week. Do these figures confirm the hypothesis that the male and female babies are born in equal number? **(6)**

Q8. (a) Write a short note on the chi-square test of goodness of fit of a random sample to a hypothetical distribution. State the Yates correction for continuity. **(6)**

(b) In a survey of 225 boys of which 85 were intelligent, 45 have educated fathers, while 95 of the unintelligent boys have uneducated fathers. Do these figures support the hypothesis that educated fathers have intelligent boys? (Given $\chi^2\ (1\ d.f., 5\%) = 3.84$ **(6)**

Q9. (a) Explain Wilcoxon Signed Rank test method with the underlying assumptions. Also write their hypothesis and give one example. **(6)**

(b) The nicotine contents of two brands of cigarettes, measured in milligrams, was found to be as follows:

Brand A:	2.1	40	63	5.4	4.8	3.7	6.1	3.3		
Brand B:	4.1	0.6	3.1	2.5	4.0	6.2	1.6	2.2	1.9	5.4

Test the hypothesis, at the 0.05 level of significance, that the average nicotine contents of the two brands are equal against the alternative that they are unequal. **(6)**

x-x-x