Use separate answer script for each group

Ph.D. Course Work Examination, 2019
Statistics
Course - 1
Research Methodology and Techniques

Time : 4 Hours **Full Marks : 80**

Questions are of value as indicated in the margin

Group – A (Computer Application) Marks : 20
Answer Question No.1 and any two from the rest.

1.a) Write the corresponding formula of the following source code:

```
$$\sum_{i=1}^k n_i^2 \leq n^2 - (k-1)(2n-k)$$
```

b) What is Float? Write the names of two floating objects.
c) What is the use of the BibTeX file (.bib)?
d) Write the syntax of the ordered list structure environment.
e) What is the use of the *verbatim* environment? $2 \times 5 = 10$

2.a) Write the source code of the following equation:

$$\sin\frac{\alpha}{2} = \pm\sqrt{\frac{1-\cos\alpha}{2}}$$

b) Write the source code of the following equation:

$$A_{m,n} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

c) How to add extra horizontal space in math mode? What is the use of \mathrm{} command? $2 + 2 + 1 = 5$

3. a) Write the names of two formats of vector graphics. Why are vector graphics preferred over raster graphics?
b) When to use the "starred" (*) variants of *figure* and *table* environments?
c) Write the names of two environments to arrange multiple figures side-by-side. $2 + 2 + 1 = 5$

P.T.O.

4. a) Write the corresponding table of the following source code:

```
\begin{tabular}{|l|l|}
 \hline
 \multicolumn{2}{|c|}{Course Sheet} \\
 \hline
 CS01 &  P. Ghosh \\ \hline
 \multirow{2}{*}{CS02} & A. Saren \\
 & C. Roy \\
 & D. Sarkar \\ \hline
\end{tabular}
```

b) What is the use of \cline{} command in the *tabular* environment?
c) Briefly explain the purposes of \ref{} and \cite{} commands?

$2 + 2 + 1 = 5$

Group – B (Full Marks : 60)

Answer **any four** questions

1. Briefly describe, with examples, the life cycle of data. 15
2. a) A certain chemical process is said to have produced 15 or less pounds of waste material for every 60 lbs. batch with a corresponding standard deviation of 5 lbs. A random sample of 100 batches gives an average of 16 lbs. of waste per batch. Test at 10 per cent level whether the average quantity of waste per batch has increased. Compute the power of the test for $\mu = 16$ lbs. If we raise the level of significance to 20 per cent, then how the power of the test for $\mu = 16$ lbs. would be affected? 9

   b) Suppose the following ten values represent random observations from a normal parent population: 2, 6, 7, 9, 5, 1, 0, 3, 5, 4. Construct a 99 per cent confidence interval for the mean of the parent population. 6

3. Distinguish between the following:

   (a) Statistic and parameter; (b) Confidence level and significance level; (c) Random sampling and non-random sampling; d) Point estimation and interval estimation and e) Simple and composite hypothesis. 5X3

4. What points will you keep in mind while preparing a research report? Explain. 15

5. R is an open source software for statistical computing and data processing. Write in brief the followings:

   a) Disadvantages of R
   b) How R can be used in predictive analysis (in brief)
   c) Distinguish between matrix, dataframe and list in R

3X5

6. a) Name any two drop and down menu softwares which may also be used for statistical computations and state their adantages and disadvantages.

   b) Write down the code/ sequence to plot scatter diagram and perform regression analysis in those two softwares and R. Plots should contain title and axes names. 5+10

7. Describe some of the important research designs used in experimental hypothesis-testing research study. 15

Ph.D. Course Work Examination, 2019

Statistics

Course – 2 (Elective)

Applied Nonparametric Techniques

Time : 4 Hours **Full Marks : 80**

Questions are of value as indicated in the margin

Answer any four questions

1. a) What is a kernel in the context of nonparametric statistics?
   b) Define U statistic.
   c) What is degenerate U statistic of order m?
   d) Write down the asymptotic distribution of U statistic.
   e) Write down the asymptotic distribution of degenerate U statistic with order 1.

   4×5=20

2. a) Discuss Kolmogorov Smirnov one sample test for goodness of fit test.
   b) Define Wilcoxon signed rank statistic and deduce its asymptotic distribution with mean and variance.
   c) Using Wilcoxon signed rank statistic perform a test of location for one sample.

   5+3+7+5=20

3. a) Define linear rank statistic on one sample location problem.
   b) Show that Wilcoxon rank statistic for one sample location problem is a linear rank statistic.
   c) Propose any nonparametric test for testing the equality of medians.
   d) Define Nadaraya-Watson estimator in the context of a locally weighted average.

   4+6+6+4=20

4. a) Define Kendall's τ.
   b) Propose an unbiased estimator of kendall's τ.
   c) Show that this unbiased estimator is consistent as well.
   d) Deduce the formula of Spearman rank correlation. 3+3+7+7=20

5. Write short notes on **any four** of the following :
   a) Mann-Whitney U test
   b) Asymptotic property of U statistic
   c) Chi-sq. Goodness of fit test
   d) Kernel Regression
   e) Nonparametric measure of association. 4×5=20

———

Ph.D. Course Work Examination, 2019

Statistics

Course – 2 (Elective)

Geostatistics

Time : 4 Hours **Full Marks : 80**

Questions are of value as indicated in the margin

(Computer Laboratory may be used, if necessary)

Answer any five questions

1. a) Differentiate between the scale and resolution for plotting spatial data. 6
   b) Describe the steps in R-code for plotting a gridded dataset over a specified region of your choice (You can choose India or some other region as well). 10
2. a) In a record of the number of times a river floods in rainy season over a period of 140 years, the following data were compiled:

No of years	No. Of floods
19	0
39	1
42	2
23	3
11	4
6	5

From the above data find the probability that there will be 2 floods in the next year. Answer the question after commenting on the distributional assumption you would made. 10

b) Following is the annual rainfall data (in cm) at two places A and B for 24 years in sequence from the initial year to the 24^{th} year.

A: 39.59, 19.93, 23.91, 29.38, 43.09, 25.34, 49.35, 39.62, 42.90, 53.30, 57.66, 37.05, 34.14, 38.01, 52.4, 32.20, 47.81, 33.98, 39.46, 37.78, 63.24, 39.04, 60.51, 38.08

B: 39.48, 17.81, 24.47, 24.32, 42.18, 23.41, 45.13,42.83, 46.94, 51.51, 57.50, 34.35, 34.29, 38.65,, 50.32, 29.94, 45.24, 34.13, 40.68, 35.54, 57.24, 42.05, 55.53, 37.45

Assumming that the above sequential data for 24 years constitute a representative sample of the rainfall at the two places can we consider the two places to have, on the overage, same amount of rainfall? 6

3. a) A random sample of 100 boring points has been selected from an area and the depth of the ground water level is recorded. The mean and standard deviation of the boring depths are 900 cm and 480 cm respectively. Find out the confidence limits for the mean depth of ground water in the whole area. 6
   b) Summarize the procedure for nearest neighbour analysis. For an area measuring 60 sq.km 16 settlements are there. If the mean of the distances of each settlement and between their nearest neighbour is calculated to be 1.45 km, find out the pattern of settlement distribution and test its significance. 10
4. a) Explain spartial auto-correlation. Comment with reasons, on the spatial auto-correlation on the following point patterns :

P.T.O.

0	0	2	0
1	3	0	1
0	2	0	2
0	2	1	2

Census A

3	2	2	0
2	2	1	0
2	1	0	0
1	0	0	0

Census B 10

b) Using Moran's auto-correlation coefficient compute and comment on the autocorrelation for a small region with the following continuity matrix 6

0	1	0	1
1	0	1	1
0	1	0	1
1	1	1	0

5. Using a package or otherwise describe and interpret, with example, how analysis of air pollution of a place can be carried out. You need to perform the followings.
   i) Plot the summary statistics for the whole dataset, two particular years, with histogram and density plots of atleast two variables say PM 2.5, NO_2, NOX etc.
   ii) Provide a calendar plot for the pollutatant O_3 for a particular year and comment.
   iii) Repeat the plot of ii) along with wind direction and wind speed
   iv) Draw seasonal plots of NO_2 vs NOx for any particular year of the available data.
   16

6. Describe different metrics of measuring forecast accuracies in case of rainfall forecast.
   16

———

Ph.D. Course Work Examination, 2019
Statistics
Course – 2 (Elective)
Statistical Ecology

Time : 4 Hours **Full Marks : 80**

Questions are of value as indicated in the margin

Answer any five questions

1. Discuss population projection through mathematical curves. Describe 'Pearl and Reed' as well as 'Rhodes' method for fitting a logistic curve. 4+(6+6)

2. Describe nearest neighbor methods and its application on ecology. If $x_1, x_2, \cdots, x_n$ are observations on r^{th} nearest individual distances, find maximum likelihood estimator $(\widehat{D})$ of the average number of individuals per unit area (D). Show that, $\widehat{D}$ is a biased estimator of D. Also obtain an unbiased estimator of D. 6+4+3+3

3. Define a life table and discuss its usage. Describe the structure of a complete life table. Briefly discuss abridge life table. (3+3)+6+4

4. What do you mean by MSY? Why it is useful in ecology? Obtain MSY for Gompertz and Logistic model for population growth with linear type harvesting. 3+3+(5+5)

5. Write short note on: *i*) Capture-recapture model *ii*) Leslie Matrix Model *iii*) Pure Birth Process. 6+6+4

6. Discuss the concept of diversity of an ecosystem. Briefly describe two statistical methods for measuring diversity. 4+(6+6)

7. Consider a two-species model as follows:
$$\frac{dN_1}{dt} = r_1 N_1\left(1 - \frac{N_1}{K_1} + a_1\frac{N_2}{K_1}\right), \quad \frac{dN_2}{dt} = r_2 N_2\left(1 - \frac{N_2}{K_2} + a_2\frac{N_1}{K_2}\right)$$
Where N_1, N_2 are the population density of species I and species II respectively at time t and a_1, a_2, K_1, K_2 are positive parameters. Non-dimensionalize the system and draw the nullclines. Hence or, otherwise determine the equilibrium points and analyze the stability of all the equilibrium points. (3+3)+(4+6)
