GLS UNIVERSITY
Bachelor of Computer Applications (BCA)
(Foundation Course)
Semester-III
210302301 STATISTICS FOR DATA ANALYSIS

1. Course Objective:

- To get working knowledge of statistical methods.
- To understand statistical analysis using R programming.
- To apply the knowledge of statistics in the field of computer science and application development.

2. Course Duration:

The course will have sessions which are divided into five modules. Each module consists of six sessions of 60 minutes each and carries a weightage of 20%.

3. Course Contents:

Module No.	Modules/Sub-Modules	No. of Sessions	Marks Weightage
I	**Statistics: Overview** ● Introduction ● Meaning of statistics ● Function of statistics ● Scope and Importance of statistics ● Limitations of Statistics ● Measure of central tendency ○ Mean ▪ Arithmetic Mean ▪ Arithmetic Mean of grouped frequency distribution ▪ Combined Arithmetic Mean ▪ Advantage, disadvantages of Arithmetic Mean ○ Median ▪ Individual frequency distribution ▪ Ungrouped frequency distribution ▪ Grouped frequency distribution ▪ Advantages, disadvantages of Median ○ Mode ▪ Individual frequency distribution ▪ Ungrouped frequency distribution ▪ Grouped frequency distribution ▪ Advantages, disadvantages of Mode	06	20%
II	**Measures of Dispersion** ● Range ○ Coefficient of Range ○ Advantages and disadvantages of Range ○ Mean Deviation ▪ Absolute M.D.	06	20%

	▪ Advantages and disadvantages of M.D. ● Quartile ○ IQR ○ Quartile Deviation ▪ Coefficient of Q.D. ▪ Advantages and disadvantages of Q.D. ● Variance and Standard Deviation ○ Coefficient of S.D. ○ Advantages and disadvantages of S.D		
III	**Correlation and Regression** ● Correlation Analysis ○ Introduction ○ Types of correlation ○ Positive, negative and zero correlation ○ Linear and non-linear correlation ○ Simple, multiple and partial correlation ○ Positive, negative and zero correlation ○ Karl Pearson method for measuring correlation ● Regression analysis ○ Introduction ○ Method of least square ○ Regression lines ○ The regression equation Y on X ○ The regression equation X on Y ○ Regression coefficients and its properties (without proof)	06	20%
IV	**Fundamentals of probability** ● Introduction ● Random experiment ● Trial event ● Favourable cases ● Equally likely events ● Mutually exclusive events ● Exhaustive events ● Dependent and independent events ● Classical approach to probability ● Statistical approach to probability ● Modern approach to probability ● Theorems of probability (without proof) ○ Addition (only for two events) ○ Multiplication (only for two events) ● Bay's Rule (only for two events)	06	20%
V	**Statistical Analysis using R programming** ● The language R and the environment ● Advantages and disadvantages of R ● R console ● R studio ● R Data types ● R Variables, constants & vectors ● R operators & conditional statements ● Expressions and assignment	06	20%

	● R input & output ● R Matrices ● Importing data from files ● Data visualization ○ Histogram ○ Scatter plot ● R Applications		

4. Teaching Methods:

The following pedagogical tools will be used to teach this course:

1. Lectures and Discussions
2. Assignments and Presentation
3. Video, e-learning

5. Evaluation:

The students will be evaluated on a continuous basis and broadly follow the scheme given below:

1.	Assignments / Quizzes, etc.	30% (Internal Assessment)
2.	Internal Examination	20% (Internal Assessment)
3.	External Examination	50% (External Assessment)

6. Basic Text Books:

Sr. No	Author/s	Name of the book	Publisher	Edition
T1	Padamalochan Hazarika	Business statistics	S. Chand	Latest

7. Reference Books:

Sr. No	Author/s	Name of the book	Publisher	Edition
R1	N G Das and J K Das	Business Mathematics and Statistics	Tata McGraw Hill Education Private Limited	Latest

8. E-resource:

Sr. No.	Link
1	http://www.biostat.jhsph.edu/~ajaffe/docs/undergradguidetoR.pdf
2	https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf
3	http://nptel.ac.in/courses/111104079/3
4	http://nptel.ac.in/courses/111104079/16
5	http://nptel.ac.in/courses/111104079/17

9. Session Plan:

Session No.	Topics / Chapters
1	Statistics: overview
2-3	Mean
4-5	Median
5-6	Mode
7-8	Range and Range Coefficient

9-10	Quartile and Quartile Deviation
11-12	Variance and Standard Deviation
13-14	Correlation Analysis
15-16	Regression Analysis: Introduction, method of least square, The regression equation Y on X
17-18	The regression equation X on Y, Regression coefficients, properties
19-20	Probability: various definitions
20-21	Classical, Statistical, modern approach to probability, theorems
22-24	Bays rule and practice
25	The language R and the environment, advantages and disadvantages of R
26-27	R console, R studio, data types, expression and assignments
28-30	Data Visualization, Demonstration

10. Learning Outcome:

On successful completion of this course unit students will be able to

- Appreciate the usefulness of computational methods in modern statistics
- Be able to apply the methodology to standard problems
- Produce and interpret numerical summary
- Have a brief idea of R programming