Title :Data Analysis and Statistics

Objectives:

The course is targeting the following objectives using a variety of modern computing tools/software/libraries like Minitab, Statistical Package for Social Sciences (SPSS), R and/or Python.

   a.   A breadth and depth of univariate & multivariate computing skills necessary to apply high level analytical thinking to real world problems. It includes an exposure to the essentials of computing tools/languages and statistical software use and, hence, the ability to adopt technology know-how demanded in the modern job market.

   b.   Knowledge of dealing strategies concerning big data, data science and data-based decision making while analysing core problems of the society considering input of quantitative and/or qualitative characteristics.

Outcomes:

After successful completion of the course, students will be able to

   a.   Learn computing tools/software of modern time
   b.   Understand and apply dealing strategies of quantitative & qualitative variables including collection, visualization, summarization and inferential
   c.   Design and conduct hypothesis tests to make inferences about population parameters
   d.   Perform regression analysis to model relationships between variables for predictions
   e.   Study sampling strategies, and principles & protocols of questionnaire designing

Course Code: CSE-883

Credit Hours: 3-0

Course Contents:

   1.   An introduction to statistical data and data sources
   2.   Data visualization: Principles & methods (univariate & multivariate, qualitative & quantitative cases)
   3.   Descriptive analysis (measures of location, dispersion & shape)
   4.   Correlation (simple, partial, multiple, point-biserial & association of attributes)
   5.   Linear & non-linear models for quantitative/qualitative dependent (output) & independent variable(s) (input), like Least Squares Regression & Logistic Regression, etc., their estimation, testing and assessment methods
   6.   Probability: Concepts and Computations, Conditional Probability, Random Variables and Probability Distribution (Both Discrete and Continuous Cases), Commonly used probability distributions like Binomial, Poisson, Normal, Lognormal, Uniform, etc.
   7.   Bayesian statistics and modelling
   8.   Estimation (Point Vs Interval estimation for mean & variance while dealing with one & two-populations)
   9.   Testing of hypothesis concerning mean & variance while dealing with one & two-populations
   10. Analysis of variance (one-way and two-way)
   11. Sampling (Probability & Non-Probability Sampling), Concepts of Questionnaire designing and validation

Course Contents with proposed contact Hours (Weekly plan):

Week	Contents coverage
1	An introduction to statistical data and data sources, Data visualization: Principles & methods (univariate & multivariate, qualitative & quantitative cases)
2	Descriptive analysis (measures of location, dispersion & shape)
3	Quiz-1 and Assignment-1 Correlation (simple, partial, multiple, point-biserial & association of attributes)
4	Linear models for quantitative dependent (output) & independent variable(s) (input), like Least Squares Regression, estimation, testing and assessment methods
5	Non-linear models for qualitative dependent (output) & independent variable(s) (input), like Logistic Regression, etc., estimation, testing and assessment methods
6	Quiz-2 and Assignment-2 Probability concepts and computations, Conditional Probability
7	Random Variables and Probability Distribution (Both Discrete and Continuous Cases)
8	Mid-term exam
9	Commonly used discrete probability distributions like Binomial, Poisson, etc.
10	Commonly used continuous probability distributions like Normal, Lognormal, Uniform, etc.
11	Bayesian statistics and modelling
12	Estimation (Point Vs Interval estimation for mean & variance while dealing with one & two-populations)
13	Testing of hypothesis concerning mean & variance while dealing with one & two-populations, Analysis of variance (one-way and two-way)
14	Quiz-3 and Assignment-3 Sampling (Probability & Non-Probability Sampling), Concepts of Questionnaire designing and validation
15	Recap and Miscellaneous items
16	Final-term exam

Details of lab work/workshop practice, if applicable:

Not applicable.

Recommended reading, including textbooks, reference books with dates

1. Introductory Statistics, 4th edition by Sheldon M. Ross. 2017. ISBN-13 978-0128043172. Publisher Academic Press
2. Applied Regression Analysis, 3rd edition by Norman R. Draper & Harry Smith. ISBN: 978-0-471-17082-2. Publisher WILEY
3. Applied Statistics and Probability for Engineers, 7th edition by Douglas C. Montgomery, George C. Runger. 2018. ISBN 978-1-119-40036-3. Publisher WILEY
4. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter by Wes Mckinney, 2nd edition 2017. ISBN-13 978-1491957660. Publisher O'Reilly Media
5. Probability and Statistics, 4th edition by Morris H. DeGroot, and Mark J. Schervish. ISBN 13: 978-0-321-50046-5. Publisher PEARSON.
6. Introduction to Statistical Theory Part-I and Part-II by Prof. Sher Muhammad Chaudhry & Dr. Shahid Kamal. Publisher Ilmi Kitab Khana

Nature of Assessments

1. Mid-term exam
2. Final-term exam
3. Quizzes
4. Assignments

Comparative Chart:

S. No.	Existing	Proposed
1	Introduction to probability	Retained
2	Statistical Models and probability distributions	Retained
3	Estimation, Error and Uncertainty	Retained
4	Hypothesis testing	Retained
5	Linear models and regression	Retained
6	Binomial and Poisson processes	Retained
7	Baysian parametric modelling	Retained
8	Bayesian modelling using Monte Carlo methods	Retained
9	Introduction to statistical/data analysis software	Deleted
10	Introduction to Visualization	Retained
11	Principles of visualization	Retained
12	Plotting and visualization techniques	Retained
13	Trade-offs in visualization	Retained
14	An introduction to statistical data and data sources	New content
15	Descriptive analysis for quantitative/qualitative variable(s) (measures of location, dispersion and shape)	Revised, Addition of New Content
16	Correlation (simple, partial, multiple, point-biserial and association of attributes)	Revised, Addition of New Content
17	logistic regression	Revised, Addition of New Content
18	Lognormal probability distribution, Uniform probability distribution	Revised, Addition of New Content
19	Analysis of variance (one-way and two-way)	Revised, Addition of New Content
20	Sampling (probability and non-probability sampling), Concepts of questionnaire designing and validation	Revised, Addition of New Content