STA 2140 EXPLORATORY DATA ANALYSIS   (45 Contact Hours)

Pre-Requisites:
None

(a) Course Purpose

To enable the student to apply various data exploratory methods using various software.

(b) Learning outcomes

By the end of this course the student should be able to.

(1) Define basic statistical concepts.

(2) Use methods of collecting data, and representing data in tabular or graphical form, using the computer.

(3) Apply methods of computing summary statistics using the computer.

(4) Apply the principles of correlation and repeated measures data using the computer.

(c) Course Description

Some basic concepts: definition of statistics. Populations and samples, randomness and independence. Data sources, data types, methods of data collection, numerical summaries of data, data exploratory techniques: data displays; box plot, stem and leaf, histogram, charts, diagrams. Frequency distribution: tables, graphical dis- plays, scatter plots, frequency graphs. Summary statistics: measures of location and dispersion, skewness and kurtosis. Repeated measures data, correlation. Use of R/GensStat/SPSS/Excel software, with real data.

(d) Course outline

Week	Content	Method
1	Basic concepts: definition of statistics, populations, samples, randomness and independence.	Theory
2	Data sources, data types, methods of data collection.	Theory
3	R programming: Fundamentals	Practical
4	R programming: Fundamentals	Practical
5	R programming: Fundamentals	Practical
6	**Cat 1**	Theory and practical
7	R programming: Data import and export	Practical
8	R programming: Data wrangling	Practical
9	Numerical summaries using R: measures of location and dispersion.	Practical
10	Data visualization in R	Practical
11	Correlation in R	Practical
12	**Cat 2**	Theory and practical

(e) Teaching Methodology

Lectures, Tutorials, Self-Reading, Discussions and Student Presentations.

(f)Instructional Material and Equipment

Black or White Boards, Chalk or White Board Markers, Dusters, Computer and Projector.

(g) Course Assessment

End of Semester Examination (70%); Continuous Assessment Tests (20%); Assign- ments (10%).

(h) Course Text Books

[1]Box G.E.P., Hunter W.G. & Hunter J.S. (1978), *Statistics forExperimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley and Sons. ISBN-13: 978-0471718130, 1978.

[2]Hastings E. & Peacock, *Statistical Distributions*, 4th. Ed., John Wiley and Sons. ISBN-13: 978-0470390634, 2000.

[3]William C., *Visualizing Data*, Hobart Press ISBN-10: 0963488406 ISBN-13: 978- 0963488404, 1993.

(i) Course Journals

[1]Statistical Methodology, ISSN:1572-3127.

[2]Statistical Methods and Applications, ISSN: 1618-2510; 1613-981X.
[3]Journal of Statistical Computation and Simulation, 0094-9655.

(j) Reference Text Books

[1]Chambers, J.W., Cleveland, B.K. & Tukey P., *Graphical Methods for Data Anal- ysis*, Wadsworth, ISBN-13: 978-0534980528, ISBN-10: 053498052X, 1993.

[2]Cleveland W. and McGill M., *Dynamic Graphics for Statistics*, Chapman and Hall/CRC, ISBN 9780534091446, 053409144X, 1988.

[3]Draper and Smith, *Applied Regression Analysis*, 2nd ed., John Wiley and Sons, ISBN 0-387-94142-8, 1981.

[4]Du Toit S., & Stumpf, *Graphical Exploratory Data Analysis*, Springer-Verlag ISBN 10: 0387963138 ISBN 13: 9780387963136, 1986.

(k) Reference Journals

[1]Communications in Statistics. Simulation and Computation, 0361-0918; 1532- 4141.

[2]Communications in Statistics. Theory and Methods, 0361-0926; 1532-415X.
[3]Journal of Modern Applied Statistical Methods, ISSN: 1538-9472.