Course Name:	MULTIVARIATE STATISTICAL ANALYSIS
Course Code:	STAT6632
# of Contact Hrs:	One Semester (13 weeks - 36 hours of lectures and 24 hours of computer lab sessions)
Credits:	4
Level:	Graduate
Pre-requisite:	None

RATIONALE

Multivariate techniques are applied to a wide array of disciplines, such as business, health sciences and economics. In many cases, a multivariate method is used as one component to better understand multi-dimensional data such as data reduction and how variables are correlated.

AIM

The aim of the course is to introduce a variety of standard statistical methods used to analyze multivariate data. Emphasis will be placed on developing the theory of these methods as well as the various interpretations of results derived from applying these methods. The (free) R statistical computing package will be used for data analyses.

LEARNING OUTCOMES

On completion of this course, the student should be able to:

- Pose informed and insightful questions concerning multivariate data and its analysis.

- Evaluate various aspects of a multivariate dataset and choose an appropriate type of analysis.

- Demonstrate proper data manipulation and exploratory data analysis skills for multivariate data.

- Perform

- Multivariate Analysis of Variance (MANOVA)

- Predictive Discriminant Analysis (PDA)

- Principal Components Analysis (PCA)

- Exploratory Factor Analysis (EFA)

- Cluster Analysis

- Write a clear report of statistical analysis methods and results.

TEACHING METHODS:

Students will be exposed to the theoretical aspects of Multivariate Statistical Analysis via lectures intertwined with tutorials. The practical components will include applying the methods developed to real-world datasets. This will be done via computer sessions that will run parallel with lectures. Both theoretical and practical homework will be given at an average of one per week. In the computer laboratory sessions they will learn to handle large data sets, parallel computing and to implement theoretical/numerical schemes learnt in the lecture hours.

There will be 36 hours of face-to-face lectures and 24 hours of supervised computer lab sessions (counted as 12 contact hours). Course material, including laboratory datasets   will be posted on the course webpage hosted at http://ourvle.mona.uwi.edu/.

SYLLABUS

Linear Algebra (Matrix Theory) Review [2 hrs lecture + 2 hrs lab]

Random vectors [1 hour]

Multivariate distributions–normal, Wishart, Hotelling's-T, Skew-T, Skew-Normal [6 hrs lecture + 4hrs lab]

Estimation and testing of multivariate distribution parameters [3 hrs lecture + 2 hrs lab]

Multivariate Analysis of Variance (MANOVA) [5 hrs lecture + 3 hrs lab]

Predictive Discriminant Analysis (PDA) [3 hrs lecture + 2 hrs lab]

Principal Components Analysis (PCA) [8 hrs lecture + 6 hrs lab]

Exploratory Factor Analysis (EFA) [5 hrs lecture + 3 hrs lab]

Cluster Analysis [3 hrs lecture + 2 hrs lab]

Total contact hrs = 48 i.e. 36 hrs lecture + 24 hrs lab

ASSESSMENT:

The course assessment has two components consisting of coursework (50%) and a final exam (50%)

Two In-course and in-class tests – 20% of overall grade;

One Laboratory assignment (project and report) – 20% of overall grade;

Two graded (at home) assignments – 10% of overall grade;

One Final exam (2 hours) – 50% of overall grade.

REFERENCE MATERIAL

Books:

Johnson, R.A. and D.W. Wichern, (2008). *Applied Multivariate Analysis, 6th Ed.*, Prentice Hall. ISBN-10: 0131877151

Sharma, S. (1996). *Applied Multivariate Techniques. John Wiley & Sons, Inc.* ISBN-10: 0471310646

ONLINE RESOURCES:

www.r-project.org: R project homepage – software downloads, tutorials and documentation;

http://archive.ics.uci.edu/ml/: Free Multivariate Datasets are available here.

http://little-book-of-r-for-multivariate-analysis.readthedocs.org/en/latest/: This site gives a great introduction to analyzing multi-dimensional datasets using R.