Course Name:	**LINEAR MODELS**
Course Code:	**STAT2003**
Number of Credits:	**3**
Semester:	**II**
Level:	**2**
Pre-requisites:	**STAT1001: Statistics for the Science, STAT2001: Inferential Statistics**

RATIONALE: Whether you work in the public or private sector and manage data, government officials or managers are increasingly responsible for giving clear answers to any questions such as: Is our school improving? Are we as a society becoming unskilled? Questions like these and more need answering. We need to turn these data, which have been collected into information, so that an informed decision can be made. The purpose of this course is to develop students' abilities in understanding and solving practical statistical problems, and to teach them how to choose appropriate techniques, analyse data and present results.

COURSE DESCRIPTION: The course will consist of a mixture of lectures and practical work (which will be assessed by the student's completion of practical assignments to be submitted); Computer practical sessions in which the statistical package R, will be used to analyse the data. This will be the basis on which continuous assessment will be conducted. The lectures will focus on statistical modelling, including selection of appropriate models, the analysis and interpretation of results and diagnostics. Exploratory and graphical techniques will be considered, as well as formal statistical procedures. Practical applications will be emphasised throughout.

LEARNING OUTCOMES: On successful completion of the course, the students should be able to:

- Construct appropriate statistical models in a wide range of application e.g. medical, biometry and environmental
- Analyse data and make inferences in a wide range of application
- Validate and assess statistical models to ensure that the models are appropriate
- Produce a written report for a non-technical individual

- Interpret computer outputs

COURSE CONTENT

1. **Exploratory Data Analysis:** numerical summaries:- mean, median, mode, trimmed mean, quartiles, range, variance, standard deviation, percentiles, skewness, kurtosis, semi-interquartile range, inter-quartile range, coefficient variation; graphical summaries:- Dotplot, Stem-and-Leaf diagram, Box-and-Whisker plot, Rootograms, Radar/Spider plots, Matrix plot; Quantile function:-theoretical distributions and empirical distributions, QQ plots; Parameter estimation: bootstrap method

2. **Linear Regression:** Median polishing technique, Resistant method for fitting straight line, Additive models:- structure and fitting, Polynomial regression;

3. **Logistic Regression:** Introduction, fitting simple model, Inferences:- confidence interval, significance testing; Multiple Logistic regression, Odds ratios, Interpretation of fitted logistic models; Assessing model: Goodness-of-fit, Pearson's chi-square statistic and deviance, diagnostic measures, validation; Case-control studies Application

4. **Analysis of Variance:** One-way and Two-way Analysis of variance with and without interaction, Additive models, Regression approach to ANOVA

TEACHING METHODOLOGY: The course will consist of a mixture of interactive lectures, computer laboratory classes and practical work. The **39** contact hours is comprised of **24 hours of lectures** and **30 hours supervised laboratory time inclusive of unsupervised hours for the projects**. There will be two data-based projects. Project 1 will be handed out at the laboratory session in week four and will be due in week six. Project 2 will be handed out at the lecture in week 10 and will be due in week 12. Students will be given a dataset taken from a real situation in which they will have to discuss in a group, analyse and individually write their own report. Course materials such as notes, problem-based learning, projects and solutions will be posted on the webpage http://ourvle.mona.uwi.edu/

ASSESSMENT: This course assessment is **100% coursework**. The coursework elements are as follows:

(i) Project 1 – **40%**
(ii) Project 2 – **40%**

(iii) Problem papers (about 2) **20%**

Even though students work in groups, each is graded individually.

REFERENCE MATERIALS:

Books:

Prescribed

1. Chambers J.M. *1983*, *Graphical Methods for Data Analysis*, *1st edition,* Springer, ISBN-13: 978-0871504135

Highly Recommended

2. Hosmer, D. and Lemeshow S. *Applied Logistic Regression*, *2nd edition*, *2000*, Wiley-Interscience Publication, *ISBN-13: 978-0471356325*

Recommended

3. Venables W.N. and Ripley B D, *2010*, *Modern Applied Statistics with S*, *4th edition*, *ISBN-13: 978-1441930088*

4. Hoaglin, D., Mosteller F. and Tukey J. *Understanding Robust and Exploratory Data Analysis*, *1st edition*, *2000*, *ISBN-13: 978-0471384915*

ONLINE RESOURCES

http://www.causeweb.org/
http://www.getstats.org.uk/
http://www.ats.ucla.edu/stat/seminars/statteach/sites.htm
http://onlinestatbook.com/rvls.html