Course Name: **REGRESSION ANALYSIS**

Course Code: **STAT3001**
Level: **3**

Semester: **1**

Number of Credits: **3**

Pre-requisites: **STAT2001 and MATH2410 (background)**

RATIONALE: Industry leaders are increasingly relying on internal and macro data to improve organisation performance. The effective use of data depends on how successful an organisation is, in applying data analysis tools, data management and modelling. The data relating to a company's activities in most cases involve elements of uncertainty. Hence, statistical and probabilistic tools have been widely used to model the collected data. Thus, in the short, medium and long-term business strategies are then developed; based on these models in order to maximise performance.

COURSE DESCRIPTION: This course introduces and examines the role of contemporary statistical methods in improving business and industrial processes. The topics to be discussed include modern statistical thinking, simple linear regression where we only look at one independent variable; Are our assumptions valid? What happens if we have more than one predictor variable? (multiple regression:-dummy variables and interaction), what do we do if we encounter a violation in our model assumptions and other associated problems? Finally, how do we choose the 'best' model. The course will rely on the use of the statistical package R.

LEARNING OUTCOMES: At the end of the course, students will be able to:
- Apply statistical concepts in a regression setting
- Apply the main properties of least square estimator
- Construct confidence intervals in a regression settings
- Demonstrate the results of statistical analysis to a non-technical audience
- Analyse and interpret statistical output from statistical software package R/Minitab
- Justify statistical hypothesis in linear models
- Assess and validate statistical models

CONTENT

- **Introduction:** Recap of the following distributions, χ^2, t and F. Expectation, variance and covariance of linear functions; Correlation and hypothesis testing of r; Principles of least squares

- **Simple Linear Regression:** Basic underlying assumptions; Notations and Model fitting by least squares; Statistical properties of least square estimators:- expectation, variance, covariance; Estimation of σ^2; Partitioning the variability of the response; Inferences:- hypothesis testing, confidence interval and prediction interval; Coefficient of determination; ANOVA and F-test for simple linear regression model; Gauss Markov Theorem(BLUE); Computer outputs (SPSS, R, Minitab); Lack of fit; Regression through the origin

- **Residual Analysis:** Residual plots, Model Assumptions(constant variance, independence, normality), outlying and influential observations

- **Multiple Regression:** Recap of matrix algebra; Model fitting by least squares; Statistical properties of least square estimators: expectation, dispersion matrix and linear combination; Inferences:– hypothesis testing and confidence interval, ANOVA, F-test for the overall model; Extra sums squares principles; Interactions; Dummy variables; Simultaneous Confidence Interval
- **Model Building**
  - **Criteria:** R^2, adjusted R^2, s and Mallow's statistic
  - **Selection:** stepwise regression, forward and backward selection
  - **Diagnostics:** leverage value, Cook's distance measure
  - **Assumptions violation remedies:** transformation, weighted least squares

- **Multi-collinearity:** correlation coefficient between $x's$, effects on least squares estimates, variance inflator factor (VIF)

TEACHING METHODOLOGY: The course will be a mixture of lectures, tutorials and computer laboratory. The 39 contact hours is consists of: 30 hours of lectures, 10 hours of supervised laboratory time (counted overall as 5 contacted hours) and 4 hours of tutorials. The tutorial will be interspersed with the lectures by having students discuss exercises, revise material and/or group presentation as needed.

Course materials such as exercises, assignments, solutions, etc., will be posted on the webpage http://ourvle.mona.uwi.edu/

ASSESSMENT: The course assessment entails a course- work component worth **40%** and a final examination worth **60%**. The **Final examination** will be **two hours** in length.

The Coursework element will consist of three parts:

    i.   A mini-project which is worth **20%** of the student's final grade

    ii.   Problems papers/lab assignment, worth **10%**

    iii.   Mid-term examination, (1 hour) worth **10%**

REFERENCE MATERIAL

Books:

[Prescribed]

1. Sheather Simon: A Modern Approach to Regression with R, 1st edition, 2009, ISBN:978-1441918727

[Recommended]

2. Chatterjee S, Hadi A: Regression by Example, 4th edition, 2006, ISBN:978-0471746966

3. Myers R: Classical and Modern Regression with Applications, 2nd edition, 2000, ISBN:978-0534380168

On-line resources:

1. http://www.causeweb.org/resources/: Supplementary material, including datasets
2. http://www.r-project.org/: software download and documentation
3. Internet Google the topics