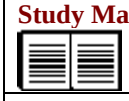
Level 6 Advanced Diploma in Data Science & Business Analytics(952) 210 Credits

Unit: Predictive Analytics with R **Exam Paper No.:** 5	**Guided Learning Hours:** 300 **Number of Credits:** 30
Prerequisites: Basic knowledge of R commands	**Corequisites:** A pass or higher in Diploma in Analytics or equivalence.
Aim: The aim of the course is to harness the power and agility of R in processing and analysing vast amounts out data. This data is increasing at exponential rate. The course provide skills required to build models using linear and logistic regression to enable better understanding and decision making in order to create value through analytic edge. R algorithm increase accuracy and calculation speed in analysing both structured and unstructured data.	
Required Materials: Recommended Learning Resources.	**Supplementary Materials:** Lecture notes and tutor extra reading recommendations.
Special Requirements: The unit requires a combination of lectures, demonstrations, discussions, and hands-on labs.	
Intended Learning Outcomes:	**Assessment Criteria:**
1. Understand how data is transforming business and how ability to effectively process data can create an analytics edge.	1.1 Define analytics. 1.2 Demonstrate R basic calculations. 1.3 Be able to create advanced data structures (Vectors and Data frames). 1.4 Be able to load and save data files. 1.5 Perform data analytics (statistics functions, scatter plots, plots and summary tables). 1.6 Be able to save script files.
2. Understand the practical implementation of basic data analysis using real life data.	2.1 Demonstrate using structure function. 2.2 Describe high-level statistical output information. 2.3 Describe purpose of data analysis functions; which.max(), names(), match() and sd(). 2.4 Describe plot(), hist(), boxplot() arguments. 2.5 Demonstrate adding variables to a data frame. 2.6 Describe table() and tapply() functions.
3. Understand the applications of linear regression in predicting/forecasting outcome variable (dependent variable).	3.1 Describe dependent *vs* independent variables. 3.2 Describe linear positive/negative relationship. 3.3 Define best fit method. 3.4 Describe regression equation. 3.5 Calculate linear regression using Least Square Method. 3.6 Calculate R-squared. 3.7 Calculate Standard Error of the Estimates (Mean Square Error). 3.8 Differentiate Sum of Square due to Error (SSE), Total Sum of Squares (SST), Sum of Squares (SST), Sum of Squares due to Regression (SSR).
4. Understand implementation of linear regression using real life data to make predictions using models.	4.1 Compute correlations. 4.2 Demonstrate making predictions through creating training and test data. 4.3 Describe coefficients. 4.4 Describe how analytics creates an edge using linear regression. 4.5 Demonstrate to check linear relationship between variables.

	4.6 Describe multicollinearity. 4.7 Define overfitting.
5. Understand application of analytics using classification through a technique called classification through a technique called logistic regression.	5.1 Describe logistic regression (binary outcome). 5.2 Compile example of questions with binary outcome. 5.3 Describe differences between linear and regression techniques. 5.4 Describe the logarithm (log) expression. 5.5 Describe multiple logistic regression. 5.6 Describe sample.plit() function. 5.7 Describe confusion/classification matrix. 5.8 Explain Receiver Operator Characteristics (ROC)
Methods of Evaluation: A 3-hour essay written paper with 5 questions, each carrying 20 marks. Candidates are required to answer all questions. Candidates also undertake project/coursework in **Predictive Analytics with R**, with a weighting of 100%.	

Recommended Learning Resources: Predictive Analytics with R

Text Books	• R for Data Analysis in easy steps - R Programming essentials by Mike McGrath. ISBN-13 : 978-1840787955 • Hands–On Programming with R by Garrett Grolemund. ISBN-13 : 978-1449359010 • R for Data Analysis by Scott McCoy. ISBN-13 : 978-1943873036
Study Manuals 	BCE produced study packs
CD ROM 	Power-point slides
Software 	R Studio