Level 5 Diploma in Foundations of Data Science Statistical Methods using Excel (951) 177 Credits

Unit: Introduction to Correlation Analysis **Exam Paper No.:** 5	**Guided Learning Hours:** 300 **Number of Credits:** 30
Prerequisites: Business terms and Excel knowledge.	**Corequisites:** A pass or higher in Diploma in Data Analytics or equivalence.
Aim: Correlation analysis is used to predict the value of a variable based on the value of another variable; between the independent (predictor) variables and the dependent (response) variable. Learners will use Excel to configure linear, multiple linear and logistic regression. The challenging tasks in analytics, especially in predictive analytics is identifying the variables or features that may be associated to the response or outcome variable of interest. The idea behind the collection data is for organisations to find answers to business questions. Finding answers to these questions, involves building predictive and prescriptive analytics models. This will lead learners to machine learning supervised and unsupervised algorithms. The purpose of the course is to create a platform for learners to grasp data science concepts in preparation for the machine learning using Python and R programs in the advanced level.	
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 the importance of linear regression, implementation and purpose of cause-and-effect studies in relation to both data science and statistics.	1.1 Explain positive and negative relationships. 1.2 Demonstrate regression positive and negative relationship equations. 1.3 Be able to calculate linear regression using Least Square Methods. 1.4 Describe Standard Error of Estimates. 1.5 Demonstrate calculating the different sum of squares (SST, SSR and SSE) both manually and in Excel. 1.6 Describe R-squared. 1.7 Be able to describe Excel regression output.
2. Understand the importance of ascertaining relationships between two variables and meaning of best fit.	2.1 Define predictor and response variables. 2.2 Be able to use and interpret a Least Squares regression linear. 2.3 Describe linear regression assumptions. 2.4 Demonstrate how to make predictions with linear regression. 2.5 Be able to conduct simple linear regression in Excel. 2.6 Explore real-life uses of linear regression.
3. Understand implementation and application of multiple predictor variables and response variable.	3.1 Describe multiple linear regression formula. 3.2 Be able to interpret multiple linear regression output. 3.3 Explain multiple linear regression assumptions. 3.4 Demonstrate conducting multiple linear regression in Excel. 3.5 Explain difference between R and R-squared. 3.6 Demonstrate how to detect multi-collinearity using Variance Inflation Factor (VIF) metric.

	3.7	Be able to create a Quantile-Quantile (Q-Q) plot.
4. Understand categorical variables when classifying observations into distinct categories and the situations when logistic regression might be used.	4.1 4.2 4.3 4.4 4.5 4.6	Explore the logistic regression equation. Be able to interpret logistic regression output. Describe logistic regression assumptions. Demonstrate implementing logistic regression in Excel. Be able to plot Receiver Operation Characteristics (ROC) curve to display sensitivity and specificity metrics. Describe real-life uses of logistic regression.
5. Understand the classification of supervised and unsupervised learning algorithms used machine learning.	5.1 5.2 5.3 5.4 5.5 5.6	Describe supervised learning algorithms. Describe the reasons behind the use of supervised learning algorithms. Describe unsupervised learning algorithms. Describe the reasons behind the use of unsupervised learning algorithms. Be able to compare and contrast supervised vs unsupervised learning algorithms. Explain the different programmes used to produce machine learning models
Methods of Evaluation: A 2½ hour essay written paper with 5 questions, each carrying 20 marks. Candidates are required to answer all questions. Candidates also undertake project/coursework in **Introduction to Linear Regression** with a weighting of 100%.		

Recommended Learning Resources: Introduction to Linear Regression

Text Books	• Introduction to Linear Regression Analysis by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. ISBN-13 : 978-1119578727 • Introduction to Linear Regression and Correlation by Allen Louis Edwards. ISBN-13 : 978-0716705628 • Linear Regression by James V Stone. ISBN-13 : 978-1916279193
Study Manuals 	BCE produced study packs
CD ROM 	Power-point slides
Software 	Excel