



Level 4 Cert in Statistics for Data Science (949) 156 Credits

Unit: Introduction to Statistics	Guided Learning Hours: 260
Exam Paper No.: 1	Number of Credits: 26
Prerequisites: Mathematical terms and Excel knowledge.	Corequisites: A pass or higher in Certificate in Excel or equivalence.
Aim: Fundamentals of descriptive statistics are mean, median, mode, and range. They are statistical tools used to summarise data sets. Learners will use learn how to use different datasets to make calculations. Learners will learn five-number summary values, namely: minimum, Q1, median (Q2), Q3, and maximum. These are used to construct a graph known as a boxplot (box and whisker plot). Interquartile Range (IQR) is used to identify outliers (extreme observations that fall outside the general scope of the other data values). The InterQuartile Range (IQR) is the difference between Q3 and Q1. So, once one finds Q1 and Q3, can calculate IQR. Q1 minus 1.5 times IQR (is the lowest it can be) and Q3 plus 1.5 times IQR (is the highest it can be). Skewness is the different shapes of a distribution, which can be normal, Unimodal, bimodal and trimodal distributions. Distribution skewed to the right (right skewed) is known as a positive skew and distribution skewed to the left (left skewed); also called Negative skew. If data that is perfectly symmetrical, the mean, median are mode are all equal. In this module, learners will also learn to create different graphs and tables.	
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: - Understand core standard statistical measures used in analysing data in order to establish clear business questions and problem goals to be achieved. - Understand how quartiles values are statistically ordered and their use in analysing data distribution, creating box-and-whisker plots visualisation and identifying outliers. - Understand how box and whisker plot is used to summarise data into five portions; namely minimum, lower quartile (Q1), median (Q2), upper quartile (Q3) and maximum; in order to highlight the dataset distribution centre, spread and any extreme outliers. - Understand the importance of skewness as a statistical measure of the asymmetry of a probability distribution around its mean and how it indicates whether the data points are evenly distributed on both sides of the average or if they lean more heavily in one direction.	Assessment Criteria: 1.1 Be able to calculate mean, median, mode and range. 1.2 Describe the advantages and disadvantages of using mean. 1.3 Be able to evaluate odd and even median. 1.4 Describe the type of data applicable to model. 1.5 Describe how range is sensitivity to outliers. 2.1 Describe Quartiles, Interquartile Range (IQR) and Outliers. 2.2 Be able to identify Q1 (first quartile) data in a range. 2.3 Be able to identify Q2 (second quartile) data in a range. 2.4 Be able to identify Q3 (third quartile) data in a range. 2.5 Demonstrate identifying if a number in a dataset is an outlier. 3.1 Identify the Five-Number summary/sections of box and whisker plot. 3.2 Demonstrate how to create Box and Whisker Plots. 3.3 Be able to calculate Interquartile Range (IQR). 3.4 Be able to calculate Q1, Q2 and Q3. 3.5 Be able to identify the presence of outliers. 4.1 Explain the use of skewness in deciphering symmetry of distributions. 4.2 Define Symmetric Distribution. 4.3 Describe Unimodal, Bimodal and Trimodal distributions. 4.4 Be able to demonstrate symmetric distribution boxplots. 4.5 Describe skewed to the Right or to the Left. 4.6 Demonstrate shapes of box plots skewed to the Left or to the Right.

5. Understand the use of different statistical graphical plots (Dot and Steam/Leaf) and Tables (Frequency and Histogram).	5.1 Demonstrate creating different statistical graphics and tables. 5.2 Describe a Dot plot. 5.3 Describe a Stem and Leaf Plot. 5.4 Be able to create a frequency table from a given set of data. 5.5 Describe the difference between histogram and bar graph.
Methods of Evaluation: A 2-hour written examination paper with Section A and Section B. Section A has 40 multiple choice questions. Section B has three essay questions, each carrying 20 marks. Candidates are required to answer all questions. Candidates also undertake project/coursework in Introduction to Statistics with a weighting of 100%.	

Recommended Learning Resources: Introduction to Statistics

Text Books 	- Statistics for Absolute Beginners by Oliver. 2nd Edition. ISBN-13 : 979-8654976123 - Statistics without Tears: An Introduction for Non-Mathematicians by Dr Derek Rowntree. ISBN-10 : 0141987499 - Statistics 101: From Data Analysis and Predictive Modeling to Measuring Distribution and Determining Probability, Your Essential Guide to Statistics by David Borman. ISBN-10 : 1507208170
Study Manuals 	BCE produced study packs
CD ROM 	Power-point slides
Software 	Excel