Programme	**LEVEL 4 EXTENDED DIPLOMA IN DATA SCIENCE**
Unit Number/ Unit Title	UNIT 5 STATISTICAL METHODS FOR DATA SCIENCE
Cohort Code:	L04SMD-U5
Unit Level	Level 4
Total GLH	Total qualification time 200/ Total Guided learning hours 90/ Self-guided learning hours 110
Credits	20 CATS/ 10 ECTS
Lecturer	
Start Date	End Date

Unit Aims	This module aims to equip students with the statistical foundations necessary for analyzing and interpreting data in the field of data science. Students will acquire the statistical foundations essential for analyzing and interpreting data within the field of data science. The module will cover key concepts such as descriptive statistics, inferential statistics, probability theory, and hypothesis testing. Students will learn how to apply statistical methods to real-world data, interpret results, and make data-driven decisions. By the end of the module, students will be proficient in using statistical techniques to extract insights from data, laying a solid groundwork for advanced data analysis and modeling.
Differentiation Strategies *(e.g. planned activities or support for individual learners according to their needs)*	The total number of students to be in the lesson is approximately 20. This is a multicultural group of students predominantly between the ages of 24 – 45, with numerous ethnic, gender, and creed background. These are UK academic level 5 students; hence it is assumed that they have practical, theoretical, or technological knowledge and understanding of a subject or field of work to find ways forward in broadly defined, complex contexts. These students must be able to generate information, evaluate, synthesise the use information from a variety of sources. Various approaches to addressing the various identified students needs will be adopted throughout the lesson. Such will include:- 1. Progressive tasks

	2. Digital resources
	3. Verbal support
	4. Variable outcomes
	5. Collaborative learning
	6. Ongoing assessment
	7. Flexible-pace learning
Equality & Diversity	Variety of teaching techniques will be employed to ensure that the needs of each individual learner are met.
Safeguarding & Prevent	Safeguarding policies and the Prevent duty are strictly observed to ensure the safety, well-being, and inclusivity of all students and staff.
Health & Safety	SIRM H&S policies will be maintained.
Learning Resources	**Teaching and Learning Materials**
	• Wasserman, L. (2013). All of Statistics: A Concise Course in Statistical Inference. Springer Science & Business Media. • Agresti, A., & Finlay, B. (2009). Statistical Methods for the Social Sciences. Pearson. • Wilks, D. S. (2011). Statistical Methods in the Atmospheric Sciences. Academic Press.

Learning Outcome	Assessment Criteria
LO1. Understand fundamental statistical concepts.	1.1 Describe basic probability theory and its application to data science. 1.2 Explain statistical distributions and their properties.
LO2. Develop skills in statistical analysis	2.1 Apply hypothesis testing and inference methods to analyze data. 2.2 Perform regression and correlation analysis.
LO3. Apply statistical methods to real-world data sets	3.1 Interpret statistical results and draw conclusions based on data analysis. 3.2 Use statistical software for data manipulation and analysis

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Introduction to Probability Theory**	**Introduction to Probability Theory** Sample spaces, events, axioms of probability	LO1: Fundamental Statistical Concepts	
2.	**Conditional Probability & Bayes' Theorem**	**Conditional Probability & Bayes' Theorem** Applications in spam filtering, medical testing	LO1: Fundamental Statistical Concepts	
3.	**Common Probability Distributions**	**Common Probability Distributions** Normal, Binomial, Poisson, and their properties	LO1: Fundamental Statistical Concepts	
4.	**Descriptive Statistics**	**Descriptive Statistics** Measures of central tendency (mean, median, mode) and dispersion (variance, SD)	LO1: Fundamental Statistical Concepts	
5.	**Statistical Visualization**	**Statistical Visualization** Histograms, box plots, Q-Q plots	LO1: Fundamental Statistical Concepts	
6.	**Sampling Methods**	**Sampling Methods** Random, stratified, cluster sampling	LO2: Statistical Analysis Techniques	
7.	**Central Limit Theorem**	**Central Limit Theorem** Implications for confidence intervals and hypothesis testing	LO2: Statistical Analysis Techniques	
8.	Half-Term Exam	- Review of LO1 topics - Practice questions and mock assessment - **Half-term assessment** based on LO1 (theory)	LO1 LO2	
9.	**Hypothesis Testing Fundamentals**	**Hypothesis Testing Fundamentals** Null vs. alternative hypotheses, p-values, significance levels	LO2: Statistical Analysis Techniques	

10.	**t-Tests & ANOVA**	**t-Tests & ANOVA** Independent samples, paired t-tests, one-way ANOVA	LO2: Statistical Analysis Techniques	
11.	**Non-Parametric Tests**	**Non-Parametric Tests** Wilcoxon, Mann-Whitney U, Kruskal-Wallis	LO2: Statistical Analysis Techniques	
12.	**Correlation Analysis**	**Correlation Analysis** Pearson's r, Spearman's rank correlation	LO3: Regression & Correlation	
13.	**Simple Linear Regression**	**Simple Linear Regression** Least squares method, R^2 interpretation	LO3: Regression & Correlation	
14.	Final Exam Preparation & Review	- Comprehensive review of all learning outcomes - Practice questions and revision of key topics		
15.	Final Exam	- **Final-term assessment** covering all learning outcomes (theory and practical elements)		
16.	Feedback & Reflection	- Review of final exam - Individual feedback on performance - Reflective discussion on key learning points		
17.	**Multiple Linear Regression**	**Multiple Linear Regression** Multicollinearity, adjusted R^2	LO3: Regression & Correlation	
18.	**Logistic Regression**	**Logistic Regression** Odds ratios, logit function	LO3: Regression & Correlation	
19.	**Model Diagnostics**	**Model Diagnostics** Residual analysis, heteroscedasticity	LO3: Regression & Correlation	

20.	**Time Series Analysis Basics**	**Time Series Analysis Basics** Trend, seasonality, moving averages	LO4: Applied Statistical Methods	
21.	**Experimental Design**	**Experimental Design** A/B testing, factorial designs	LO4: Applied Statistical Methods	
22.	**Bootstrapping & Resampling**	**Bootstrapping & Resampling** Confidence intervals, permutation tests	LO4: Applied Statistical Methods	
23.	Half-Term Exam	**Capstone Project** Full statistical analysis of a real dataset		
24.	**Bayesian Statistics Introduction**	**Bayesian Statistics Introduction** Priors, posteriors, Markov Chain Monte Carlo (MCMC)	LO4: Applied Statistical Methods	
25.	**Statistical Power & Sample Size**	**Statistical Power & Sample Size** Effect size, power calculations	LO4: Applied Statistical Methods	
26.	**Statistical Programming with Python**	**Statistical Programming with Python** Pandas for EDA, SciPy for hypothesis testing	LO5: Software & Real-World Applications	
27.	**R for Statistical Analysis**	**R for Statistical Analysis** Comparison with Python, basic R commands	LO5: Software & Real-World Applications	
28.	**Case Study: Marketing Analytics** and **Case Study: Healthcare Data**	**Case Study: Marketing Analytics** Customer segmentation using clustering **Case Study: Healthcare Data** Survival analysis, Kaplan-Meier curves	LO5: Software & Real-World Applications	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4	LO1, LO2, LO3, LO4	
30.	Final Exam		LO1, LO2, LO3, LO4	