APPROVED

DATA C9002: Statistics

Module Details

Module Code:	DATA C9002
Full Title:	Statistics APPROVED
Valid From::	Semester 2 - 2023/24 (January 2024)
Language of Instruction:	English
Duration:	2 Semesters
Credits::	10
Module Owner::	Siobhan Connolly
Departments:	Computing Science & Mathematics
Module Description:	This module allows the learners to build on fundamental knowledge in statistics and probability and use these techniques as foundation to build further knowledge in the area. This module will encompass basic descriptive statistics, probability, probability distributions, statistical inference (parametric & non-parametric),and introduce generalised linear models. The module will also discuss Frequentist statistical approaches to point estimation and interval estimation and will introduce Bayesian statistics as another alternative approach.

Module Learning Outcome	
On successful completion of this module the learner will be able to:	
#	**Module Learning Outcome Description**
MLO1	Conduct exploratory data analysis by applying fundamental concepts and techniques.
MLO2	Assess fundamental probability laws and choose the appropriate probability distribution to model given problems.
MLO3	Conduct statistical inference to allow the learner to develop hypothesis to assess, examine appropriate models, interpret results and communicate findings
MLO4	Hypothesise and examine relationships between numerical variables through correlation, simple and multiple linear regression models
MLO5	Fit and interpret generalised linear models
MLO6	Distinguish between frequentist and Bayesian statistical approaches and be able to employ Bayes' Theorem to find posterior probabilities.

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named DkIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed

Module Indicative Content
Descriptive Statistics Contingency tables; Relative Risk; Odds Ratio; measures of central tendency & variation; Graphical representation of data
Probability Theory and Probability distributions Basic Laws of Probability, Probabilistic Problem Solving, Bayes Theorem, Binomial, Poisson, Normal and other distributions
Statistical Parameter Estimation Point estimation using method of moments and maximum likelihood estimators. Consistency, bias and mean squared error.
Hypothesis Testing Both Parametric and Non-parametric methods to do one sample, two sample, paired sample and more than two samples hypothesis testing
Regression Analysis Scatterplots, Correlation, Simple Linear Regression
Generalised Linear Models Multiple Linear Regression, Logistic Regression, Loglinear Regression
Introduction to Bayesian Statistics Prior probability, Posterior probability, Bayesian inference

Module Assessment

Assessment Breakdown	%
Course Work	25.00%
Project	25.00%
Final Examination	50.00%

Module Special Regulation

Assessments

Full-time

Course Work

Assessment Type	Continuous Assessment	**% of Total Mark**	25
Marks Out Of	100	**Pass Mark**	40
Timing	n/a	**Learning Outcome**	1,2,3,4
Duration in minutes	0		

Assessment Description
Ongoing short individual & group exercises or quizzes.

Project

Assessment Type	Project	**% of Total Mark**	25
Marks Out Of	100	**Pass Mark**	40
Timing	End-of-Semester	**Learning Outcome**	1,2,3,4,5
Duration in minutes	0		

Assessment Description
Data Project 1: Practical Implementation of some aspects of content of this module that involve analysing data. This will form part of a joint project with Programming for Data Analytics

No Practical

Final Examination

Assessment Type	Formal Exam	**% of Total Mark**	50
Marks Out Of	100	**Pass Mark**	40
Timing	End-of-Semester	**Learning Outcome**	2,3,4,5,6
Duration in minutes	0		

Assessment Description
End-of-Module Final Examination

Part-time

Course Work

Assessment Type	Continuous Assessment	**% of Total Mark**	25
Marks Out Of	100	**Pass Mark**	40
Timing	Every Second Week	**Learning Outcome**	1,2,3,4,5
Duration in minutes	0		

Assessment Description
Bi-Weekly short individual & group exercises or quizzes.

Project

Assessment Type	Project	**% of Total Mark**	25
Marks Out Of	100	**Pass Mark**	40
Timing	End-of-Semester	**Learning Outcome**	1,2,3,4,5,6
Duration in minutes	0		

Assessment Description
Data Project 1: Practical Implementation of some aspects of content of this module that involve analysing data. This will form part of a joint project with Programming for Data Analytics.

No Practical

Final Examination

Assessment Type	Formal Exam	**% of Total Mark**	50
Marks Out Of	100	**Pass Mark**	40
Timing	End of Year	**Learning Outcome**	1,2,3,4,5,6
Duration in minutes	0		

Assessment Description
End-of-Module Final Examination.

Reassessment Requirement

A repeat examination
Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.

DKIT reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full-time

Workload Type	*Contact Type*	*Workload Description*	*Frequency*	*Average Weekly Learner Workload*	*Hours*
Lecture	Contact	Three one hour lectures in the week (over 1 semester). This contact may occur face-to-face or online (using an appropriate mix of synchronous and asynchronous delivery) to ensure flexibility for learners.	Every Week	3.00	3
Practical	Contact	This will be practical/tutorial 2 hours a week (over one semester).This contact may occur face-to-face or online (using an appropriate mix of synchronous and asynchronous delivery) to ensure flexibility for learners implement and discuss material.	Every Week	2.00	2
Directed Reading	Non Contact	-	Every Week	3.00	3
Independent Study	Non Contact	-	Every Week	8.00	8
				Total Weekly Learner Workload	16.00
				Total Weekly Contact Hours	5.00

Workload: Part-time

Workload Type	*Contact Type*	*Workload Description*	*Frequency*	*Average Weekly Learner Workload*	*Hours*
Lecture	Contact	1.5 lecture-hours in the week (delivered over 2 semesters). This contact may occur face-to-face or online (using an appropriate mix of synchronous and asynchronous delivery) to ensure flexibility for learners.	Every Week	1.50	1.5
Practical	Contact	This will be practical/tutorial 1 hour a week (over 2 semesters).This contact may occur face-to-face or online (using an appropriate mix of synchronous and asynchronous delivery) to ensure flexibility for learners implement and discuss material.	Every Week	1.00	1
Directed Reading	Non Contact	-	Every Week	1.50	1.5
Independent Study	Non Contact	-	Every Week	4.00	4
				Total Weekly Learner Workload	8.00
				Total Weekly Contact Hours	2.50

Module Resources

Supplementary Book Resources

Roger Berger, George Casella. (2001), Statistical Inference, 2nd.

Bradley Efron, Trevor Hastie. (2021), Computer Age Statistical Inference: Algorithms, Evidence, and Data Science, Student Edition.

David Freedman, Robert Pisani, Roger Purves. (2014), Statistics.

Jeremy Arkes. (2023), Regression Analysis: A Practical Introduction, 2nd.

This module does not have any article/paper resources

Other Resources

Website, Online Statistics Education: An Interactive Multimedia Course of Study,
http://onlinestatbook.com/

Website, Khan Academy,
http://www.khanacademy.org

Website, Data Camp,
https://www.datacamp.com/