Module Specification

Data Science

Version: 2023-24, v2.0, 27 Mar 2023

Contents

Part 1: Information

Module title: Data Science

Module code: UFMFHH-30-M

Level: Level 7

For implementation from: 2023-24

UWE credit rating: 30

ECTS credit rating: 15

Faculty: Faculty of Environment & Technology

Department: FET Dept of Engineering Design & Mathematics

Partner institutions: None

Field: Engineering, Design and Mathematics

Module type: Module

Pre-requisites: None

Excluded combinations: None

Co-requisites: None

Continuing professional development: No

Professional, statutory or regulatory body requirements: None

Part 2: Description

Overview: Not applicable

Features: Not applicable

Educational aims: This module will concentrate on developing specific transferable skills:

Scripting using R

Automating code

Understanding the properties of statistical models using simulation based approaches

Automating production of reports using the skills of reproducible research by using tools such as RStudio and knitR

It is a highly practical module, focusing on the application of a wide variety of modern statistical methods. Students are expected to engage with literature and to understand how to test the robustness of different approaches through appropriate use of simulation.

Outline syllabus: General Themes of the Module:

Applied Simulation Based Inference:

Using appropriate simulation based approaches to enable appropriate estimation of parameters in a variety of statistical models. Furthermore, simulation approaches to understand the underlying behaviour of parameters in models will be embedded throughout the entire module.

Computationally intensive approaches such as bootstrapping and randomization methods will be studied and sampling distributions empirically derived through simulation approaches (e.g. F, chi-square, etc.).

Reading Statistical Literature:

A strand throughout the entire module would be focused on improving students' understanding of underpinning statistical theory by promoting the reading and discussion of key pieces of statistical literature describing the techniques used in the rest of the module.

Core Statistical Computing:

Techniques such as Monte Carlo integration and Importance Sampling are core to much of modern implementations of statistical modelling, and as such will be an important part of the module. Different methods for developing random number generators, inverse distribution function, rejection method, etc. will also inform this part of the module.

Specific statistical techniques/approaches such as:

Dimension Reduction and Variable Selection Approaches including:

Fourier transforms

Wavelet decomposition

Boruta feature selection

Spatial Statistics with environmental applications

Clustering and Classification techniques (Supervised and Unsupervised Learning):

Unsupervised learning (clustering)

Supervised learning

Random Forests (of trees)

Semi-supervised classification (Machine-learning style statistics)

Model Selection/Validation Approaches including:

Over-sampling;

Under-sampling;

Bagging;

Cross-validation;

Double cross-validation;

will be covered throughout the module.

Part 3: Teaching and learning methods

Teaching and learning methods: Contact hours will be a mixture of computer laboratory practicals and seminar style sessions. One third of the contact time will be spent on critical appraisal of statistical methods posed by current and historic literature. Use of flipped learning approaches will feature at appropriate stages during the module.

Module Learning outcomes: On successful completion of this module students will achieve the following learning outcomes.

MO1 Plan, design and create reports of statistical results in a fully reproducible fashion

MO2 Formulate and appraise strategies to interrogate and evaluate the properties of statistical estimators

MO3 Apply sound theoretical knowledge to the selection, assessment and application of modern statistical techniques

MO4 Use judgement and subject knowledge to assess limitations of inference from statistical models based on model evaluation techniques

MO5 Appraise, interpret and explain statistical models in different contexts to both expert and non-expert audiences

MO6 Justify the appropriateness, efficiency and validity of their solution method

MO7 Demonstrate awareness of the impact of the subject on society

Hours to be allocated: 300

Contact hours:

Independent study/self-guided study = 228 hours

Face-to-face learning = 72 hours

Total = 300

Reading list: The reading list for this module can be accessed at readinglists.uwe.ac.uk via the following link https://uwe.rl.talis.com/index.html

Part 4: Assessment

Assessment strategy: The coursework will allow students to demonstrate their skills in producing a piece of reproducible research. This will allow students to comprehensively investigate the properties of statistical techniques using simulation based approaches. Students will be expected to develop their own topics for research – feedback on potential topics will be provided during group tutorials. The single written report, excluding embedded tables, figures and code, will have a word limit of 2500 words.

The practical examination will provide students an opportunity to do some limited coding under controlled conditions, with the emphasis on producing results quickly. The examination scenario(s) will come from a list of potential scenarios that will be provided to students in advance, so students will have the opportunity to consider different analysis strategies in advance of the practical examination. Having this piece of practical work assessed under controlled conditions will minimise the opportunities for plagiarism. This will consist of a 2.5 hour “coding period” and a 30 minute session consisting of a presentation and a period for questioning on the work submitted during this coding period. The 30 minute presentation will occur during the examination period, at least a day after the coding practical exam. The presentation will assess the students’ ability to explain and justify the approaches taken in the code submitted in A1, having had the opportunity to reflect on their work.

Assessment tasks:

Examination (First Sit)
Description: Practical examination coding period (2.5 hours)
Weighting: 42 %
Final assessment: No
Group work: No
Learning outcomes tested: MO3, MO4, MO7

Report (First Sit)
Description: Report
Weighting: 50 %
Final assessment: No
Group work: No
Learning outcomes tested: MO1, MO2, MO3, MO4, MO5, MO7

Examination (First Sit)
Description: Practical examination: presentation (30 minutes)
Weighting: 8 %
Final assessment: Yes
Group work: No
Learning outcomes tested: MO5, MO6, MO7

Examination (Resit)
Description: Practical examination coding period (2.5 hours)
Weighting: 42 %
Final assessment: No
Group work: No
Learning outcomes tested: MO3, MO4, MO7

Report (Resit)
Description: Report
Weighting: 50 %
Final assessment: No

Group work: No

Learning outcomes tested: MO1, MO2, MO3, MO4, MO5, MO6, MO7

Examination (Resit)

Description: Practical examination: presentation (30 mins)

Weighting: 8 %

Final assessment: Yes

Group work: No

Learning outcomes tested: MO5, MO6, MO7

Part 5: Contributes towards

This module contributes towards the following programmes of study:

Mathematics [Sep][FT][Frenchay][4yrs] - Not Running MMath 2020-21

Mathematics [Sep][SW][Frenchay][5yrs] MMath 2019-20