2024 / 25

School of Science and Computing
+353 (0)51 302037
Eleanor.Reade@setu.ie
www.wit.ie/schools/science_computing

Ollscoil Teicneolaíochta an Oirdheiscirt

South East Technological University

Module Descriptor

Data Analysis 1
(Computing and Mathematics)

Data Analysis 1 (A35482)

Short Title: Data Analysis 1
Department: Computing and Mathematics
Credits: 5 **Level:** Advanced

Description of Module / Aims

This module will introduce the student to statistical techniques in data analysis, with a particular focus on linear models. Statistical software (such as Python or R) will be used in the application of techniques studied.

Programmes

	stage/semester/status
HDip in Science in Data Analytics (WD_KDAAN_G)	1 / 2 / M

Indicative Content

- Software basics: Data structures, control structures, functions, data formats
- Exploratory data analysis: Data wrangling, visualisation and reporting
- Multiple regression: Assumptions, model accuracy, diagnostics, interpreting coefficients. Use of software to estimate regression models
- Generalised linear models: Model specification, model fitting, Analysis of deviance, inference for model parameters, model diagnostics. Applications of model methods using software

Learning Outcomes

On successful completion of this module, a student will be able to:

1. Utilise standard statistical software to perform typical exploratory data analysis, data wrangling, analysis and reporting.
2. Utilise the concepts of multiple regression to estimate regression models using appropriate software.
3. Utilise the concepts of a generalized linear model (GLM) to apply them to real data using appropriate software.
4. Establish and apply corrections to problems when applying generalized linear models to real data.

Learning and Teaching Methods

- This module is focused on data analysis and developing good practices in dealing with data – in all stages, from efficient data storage, using appropriate techniques for EDA and subsequent visualisation and reporting. As a result the emphasis is on practice with real data, with its inherent issues such as missing values, outliers, etc. Building on the Introduction to Applied Statistics module, students will learn how to model and analyse real data.

Learning Modes

Learning Type	F/T Hours	P/T Hours
Lecture	24	24
Practical	24	24
Independent Learning	87	87

Assessment Methods

	Weighting	Outcomes Assessed
Final Written Examination	50%	2,3,4
Continuous Assessment	50%	
Practical	25%	1,2,4
Practical	25%	1,3,4

Assessment Criteria

<40%: The learner demonstrates little or no understanding of the fundamentals of the learning outcomes. Displays a lack of competence with regards to associated data analysis skills and statistics.

40%–59%: The learner demonstrates limited understanding of the fundamentals of the learning outcomes. Displays some ability with regards to associated data analysis skills and statistics.

60%–69%: As well as the above, the learner demonstrates reasonable understanding of the fundamentals of the learning outcomes. Displays reasonable competence with regards to associated data analysis skills and statistics.

70%–100%: All of the above to excellent levels. In addition, the learner demonstrates comprehensive knowledge and understanding of all learning outcomes. Displays advanced competence and independence with regards to associated data analysis skills and statistics.

Essential Material(s)

- Dobson, A.J. and A.G. Barnett. *An Introduction to Generalized Linear Models*. 4th ed.. Boca Raton: Chapman and Hall/CRC, 2018.

Supplementary Material(s)

- *Introductory Statistics with R*. 2nd ed.. New York: Springer , 2008.
- "Statistical Thinking for the 21st Century: An open source textbook for statistics, with companions for R and Python." https://statsthinking21.org
- Grolemund, G. and H. Wickham. *R for Data Science*. 1st. New York: O'Reilly Media, 2017.
- Rstudio Team , A. *RStudio: Integrated Development for R*. Boston: PBC, 2020.

Requested Resources

- Room Type: Computer Lab