

CS1D6-15 Introduction to data and statistics

23/24

Department

Computer Science

Level

Undergraduate Level 1

Module leader

Khalil Challita

Credit value

15

Module duration

1 week

Assessment

100% coursework

Study location

University of Warwick main campus, Coventry

Description

Introductory description

You cannot register for this module unless you are enrolled on the BSc Computer Science and Technology Solutions Degree Apprenticeship. It is not possible to request this module as an unusual option. If you are studying at Warwick as a visiting student from overseas it is not possible to register for this module.

This module will provide students with an introduction to data and statistics, focusing on key concepts and issues in the field, the application of a range of descriptive and predictive techniques, and the appropriate use of visualisation and reporting techniques. Students will also learn basic probability, conditional probability and Bayes theorem, sampling from univariate distributions, concepts of multivariate analysis, and linear regression, applying this knowledge to identify and support data analysis requirements within their own areas of work.

Module aims

The module will introduce students to basic concepts and issues of data and handling data sets. Students will learn and apply basic statistical methods for descriptive analysis and be able to provide appropriate visualisations. They will understand and be able to apply a number of different

sampling techniques and to apply inferential methods. They will be able to identify the range of data sets, tools, and environments relevant to their own professional practice and to apply and report on appropriate techniques for basic analysis. Students will be aware of the limitations of traditional analysis techniques and the benefits of machine learning approaches.

Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

In the module students will learn:

- Basic concepts and challenges of data and information, societal impacts, and professional and legal issues.
- Descriptive techniques for data analysis
- Visualisation and reporting techniques
- Basic probability, sampling, standard discrete and continuous distributions
- Methods for sampling from univariate distributions
- An introduction to multivariate analysis
- Linear regression, Bayes Theorem, and conditional probability
- The limitations of traditional approaches and the motivation for machine learning

Learning outcomes

By the end of the module, students should be able to:

- Understand key concepts and challenges relating to data and information.
- Understand the role of computer-based storage and analysis in society and appreciate relevant professional and legal issues of ethics and security.
- Understand the life cycle stages in a data analysis project.
- Understand and be able to select and apply an appropriate range of visualisation and reporting techniques to data sets relevant to their workplace.
- Understand and be able to select and apply an appropriate range of descriptive techniques to data sets relevant to their workplace.
- Understand the need for sampling and understand basic concepts of probability.
- Understand a range of standard discrete and continuous distributions and be able to apply generic methods for sampling from univariate distributions.
- Understand the concepts of dependent variables, controls, and multivariate analysis.
- Apply techniques for multivariate analysis.
- Understand the approach of linear regression and be able to identify and perform appropriate applications.
- Understand basic concepts of conditional probability (Bayes theorem) and the basic approach to Bayesian linear analysis.
- Apply knowledge of these areas in their own area of work.
- Describe the strengths and weaknesses of traditional statistical approaches and the differences and benefits of machine learning approaches.

Indicative reading list

Schutt, R., and O'Neil, C., "Doing data science: Straight talk from the frontline", O'Reilly Media, Inc. (2013)

Tufte, ER, Goeler, NH, & Benson, R., "Envisioning information", Vol. 126, Chesire, CT: Graphics Press (1990)

Kachigan, SK, "Statistical analysis: An interdisciplinary introduction to univariate & multivariate methods", Radius Press (1986)

Subject specific skills

- Able to manage data effectively and undertake data analysis
- Import, cleanse, transform, and validate data with the purpose of understanding or making conclusions from the data for business decision making purposes
- Perform routine statistical analyses and ad-hoc queries
- Use a range of analytical techniques such as data mining, time series forecasting and modelling techniques to identify and predict trends and patterns in data
- Report on conclusions gained from analysing data using a range of statistical software tools
- Summarise and present results to a range of stakeholders making recommendations
- The quality issues that can arise with data and how to avoid and/or resolve these
- How to use and apply industry standard tools and methods for data analysis

Transferable skills

- Have demonstrated that they have mastered basic business disciplines, ethics and courtesies, demonstrating timeliness and focus when faced with distractions and the ability to complete tasks to a deadline with high quality.
 - Flexible attitude
 - Ability to perform under pressure
 - A thorough approach to work
 - Logical thinking and creative approach to problem solving
-

Study

Study time

Type	Required
Lectures	15 sessions of 1 hour (10%)
Tutorials	14 sessions of 1 hour (9%)
Practical classes	9 sessions of 2 hours 30 minutes (15%)
Work-based learning	197 sessions of 30 minutes (65%)
Total	150 hours

Private study description

No private study requirements defined for this module.

Costs

No further costs have been identified for this module.

Assessment

You do not need to pass all assessment components to pass the module.

Assessment group A

	Weighting	Study time	Eligible for self-certification
Assessment component			
Worksheets - Set 1	40%		Yes (extension)
These problems cover the following topics: Introduction to Python, Descriptive Statistics, Data Visualisation, and Probability.			
Reassessment component is the same			
Assessment component			
Worksheets - Set 2	60%		Yes (extension)
These problems cover the following topics: Preprocessing, Clustering, PCA, Linear Regression, and Sampling.			
Reassessment component is the same			

Feedback on assessment

Written and verbal

Availability

Courses

This module is Core for:

- BSc Computer Science and Technology Solutions (Data Analyst)