

STAT6114: Data Science Fundamentals

Module Outline 2018/2019

Semester 1

Contents

1. Essential Information
2. Course content
 - a) Aims of the Module
 - b) Learning Outcomes
 - c) Key Skills
 - d) Recommended Reading
 - e) Blackboard
 - f) Outline timetable
3. Assessment and Feedback
 - a) Assessment methods
 - b) Resit arrangements
 - c) Past exam papers
 - d) Feedback
4. Grade Descriptors and Marking Criteria
5. Academic Integrity and Referencing
6. Support and Troubleshooting

This module outline should be read in conjunction with the Blackboard website for the module and the Degree Handbook for your degree programme. Degree Handbooks are available here:
<https://www.southampton.ac.uk/studentservices/academic-life/faculty-handbooks.page>

1. Essential information

Staff:

Lecturer: Paul Smith (with assistance from Andy Hinde and Martin Hinsch)
Room: 58/4139
Telephone: 02380 594857 (internal x24857)
Email(s): p.a.smith@soton.ac.uk
Office Hours: available in breaks during MOffStat/MDataGov module weeks
Administrator: socstats@soton.ac.uk
+44(0)23 8059 2250
Building 2, Room 2040

Times and Dates:

Lectures: 19 November 2018 – 23 November 2018
Social Statistics Research Centre
Seminar Room, Building 39, Highfield
campus, University of Southampton

Assignments and Assessments

This module is assessed 100% by coursework, which will be issued on the last day of teaching. Since I will not mark it during the Christmas holidays, it is due in on **31 December 2018**. However, you are strongly encouraged to submit by **21 December**, as there will not be any support available after this time, either for any problems with the submission system or with the module content, as the University will be closed (and you'll be able to have a (more) relaxing Christmas).

2. Course content

a) Aims of the Module

To introduce the core data science concepts, including:

- Understanding of the different types of data sources available across government (admin data, survey data, open data, big data, etc);
- How to collect data, including innovative data collection methods, e.g. web scraping;
- Understanding the challenges with unstructured data;
- How to deal with different data types in processing and analysis;
- How to undertake basic data analysis with structured and unstructured data;
- How to present data through basic data visualisations.

b) Learning Outcomes

Having successfully completed this module, you will be able to:

- search out and critically appraise the different types of data available for a particular task.
- write and use a simple web-scraping program.
- analyse simple unstructured and linked datasets, and understand the challenges and risks of using standard analysis with these types of data.
- describe the approaches to storage and processing (including for statistical analysis) of large, unstructured datasets.

c) Key Skills

You will develop skills in:

- Supervised learning
- Unsupervised learning
- Visualisation

d) Recommended Reading

Key text

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An Introduction to Statistical Learning with Applications in R*. Available for download free from <https://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf>

The module practical exercises will be undertaken using R. So please revise the basics of R use from the Statistical Programming module, and/or familiarise yourself with the basics of R use in suitable on-line learning.

e) Blackboard

When registered for the module, you should be enrolled automatically on the module's Blackboard course and you can log on at: <http://blackboard.soton.ac.uk/>. If you do not have access to the site please let the module coordinator know.

The site contains all the relevant course materials. A printed copy of the course materials will also be provided.

You should check in regularly to ensure you see all announcements and course materials.

f) Outline timetable

The timetable is subject to day-to-day revisions (possibly more so than usual as this is the first time this module has been delivered).

Monday 19 November

10:00 – 10.30	Welcome & Introduction
10:30 - 11:15	Data analysis and visualisation
Break	
11:45 - 13:00	Visualisation principles
Break (lunch)	
14:00 - 15:20	Practical 1
Break	
15:40 - 17:00	Data sources, unstructured data

Tuesday 20 November

09:30 - 09:45	Review of Day 1
09:45 - 11:15	Supervised learning – regression 1
Break	
11:45 - 13:00	Practical 2
Break (lunch)	
14:00 - 15:30	Supervised learning – regression 2
Break	
15:30 - 17.00	Nonlinear regression

Wednesday 21 November

09:30 - 10:00	Review of Day 2
---------------	-----------------

10:00 - 11:00 Practical 3
Break
11:30 - 13:00 Classification 1
Break (lunch)
14:00 - 15:30 Practical 4
Break
16:00 - 17:00 Classification 2

Thursday 22 November

09:30 - 09:45 Review of Day 3
09:45 - 11:00 Bagging and boosting
Break
11:30 - 13:00 Unsupervised learning 1
Break (lunch)
14:00 - 15:15 Unsupervised learning 2
Break
15:45 - 17:00 Practical 5

Friday 23 November

09:30 - 09:45 Review of Day 4
09:45 - 11:00 Data collection
Break
11:30 - 12:20 Practical 6
12:20 - 12:50 Coursework assignment
12:50 - 13:00 Course wrap-up
End of taught course

3. Assessment and Feedback

a) Assessment methods

The intended learning outcomes for the module will be assessed by coursework.

b) Resit arrangements

For students not attaining the required mark there will be resit coursework in the supplementary exam period in August 2019.

c) Feedback

Formative and summative feedback are provided in the following ways:

- Informal verbal feedback will be given during lectures and tutorials for individual and group work. (You'll need to contribute regularly to group discussions to make the best use of this.)
- Informal written and verbal feedback are often provided by email or during office hours when we respond to queries about assessments, for example.
- Written feedback will be given on your assessed coursework, available via Blackboard. As per Faculty policy our aim is to get coursework back to students within 4 working weeks of submission. Bear in mind that if you hand in work late, your feedback may be delayed.
- Exam results are published only as a grade. Although individual feedback on examinations is not normally given, feedback on the strengths and weaknesses of the performance of the whole group which took an examination may be available via Blackboard.
- Students are entitled to view their examination scripts on request, your Student Office can advise on the process to be followed. You are only permitted to view an examination script to enable you to see how you can improve your future performance and no mark or other annotation on the script is negotiable or open to alteration. The absence of annotation on a script does not mean that it has not been marked.
- Feedback works two ways – we want to hear from you about any concerns you have and suggestions about how to improve modules. We do this through informal mid semester feedback, which can sometimes be used to make immediate improvements in module delivery, and through a formal questionnaire at the end of the module, which will benefit students taking it in subsequent years. In addition to these, informal feedback from you on how we are doing and what we could do better is welcome anytime.
- For further information about how your work is marked and moderated, university quality assurance processes etc., please visit the marking and feedback section in the University's quality handbook: https://www.southampton.ac.uk/quality/assessment/framework/marking_and_feedback.page?

For the feedback to be effective, it is important that you work with the feedback given and identify how you can improve your work in the future. Should you need further information about your work, get in touch with whoever marked the work.

4. Grade Descriptors and Marking Criteria

Social Statistics and Demography follow the standard the University grade descriptors available here:

http://www.southampton.ac.uk/quality/assessment/framework/principles_and_definitions.page#assessment_descriptors when marking assessed work. The marking criteria and/or marking rubric for each individual piece of assessment on this module will be made available on Blackboard and with the instructions for each assessment. Note that the rating given for each criterion is descriptive and does not necessarily relate in a direct numerical way to the mark achieved.

5. Academic Integrity and Referencing

The University places the highest importance on the maintenance of academic integrity and expects that all students will familiarise themselves with the Regulations Governing Academic Integrity available at:

<http://www.calendar.soton.ac.uk/sectionIV/academic-integrity-regs.html>

Procedures will be invoked to investigate suspected breaches of academic integrity when concerns are raised during the marking process or in connection with suspected cheating in examinations. We are aware that students may have experienced differing standards at other institutions (including those overseas) but it is essential that you take steps to ensure your full understanding of the standards expected at Southampton as significant penalties can be imposed if these standards are breached.

It is often helpful to discuss ideas and approaches to your work with your peers, and this is a good way to help you think through your own views. However, work submitted for assessment should always be entirely your own, except where clearly specified otherwise in the instructions for the assignment. In some instances working in groups will be required, and there may be occasions when work is submitted from the whole group rather than individuals. In these instances the instructions will make it clear how individual contributions to the joint work should be identified and will be assessed. If you are in any doubt, check with the person setting the assignment. If you have worked with others you should make sure that you acknowledge this in any declaration you make.

Please note that you are NOT permitted to discuss the assignment or to show any other student your written work or computer programmes or outputs.

Copying includes using another student's computer program, output or graphics.

A very useful set of interactive guides is available at <http://library.soton.ac.uk/sash/what-is-academic-integrity>. These aim to help you gain a better understanding of academic integrity and develop your skills so that your assessed work does not accidentally plagiarise the work of others.

Referencing

There are many styles of referencing used in academic publications. Unless otherwise specified the style known as the Harvard system is preferred in our Faculty. Details about how to use the Harvard referencing system can be found through the following Hartley library link:

<http://library.soton.ac.uk/sash/referencing> or by downloading the guide from: http://library.soton.ac.uk/ld.php?content_id=4660789.

If in doubt about what is required in any particular assignment, what referencing styles are appropriate etc., always ask. Your tutor or module coordinator will be able to point you in the direction of appropriate sources of advice and information.

Unfortunately, Academic integrity breaches sometimes occur. The regulations distinguish between two types of breaches of academic integrity: minor (first-time offences, "committed through inexperience or lack of understanding and ... limited in scope or their effect"), and major. The minor breaches are dealt with by individual markers, through the regular feedback process. However, everything that is not a minor breach, including all repeated cases, is a major one.

The major breaches are dealt with either by the Faculty Academic Integrity Officer or by an Academic Integrity panel, depending on the severity of the alleged breach. The outcomes from this process can vary with the maximum penalty that can be given the termination of the programme – **so please treat Academic Integrity seriously.**

6. Support and Troubleshooting

If you find yourself experiencing any study skills difficulties contact the Academic Skills Hub, level 2 in the Hartley Library, Monday - Friday: 10:00 – 12:00 & 14:00 – 16:00. <http://library.soton.ac.uk/sash>

You can also access specialized study support from Enabling Services: https://www.southampton.ac.uk/edusupport/study_support/index.page If you experience any specific difficulties with the content of the module, please contact the module coordinator.

If you are not satisfied with the response contact your Personal Academic Tutor or the Programme Coordinators (Paul SMith p.a.smith@soton.ac.uk and Angela Luna Hernandez A.Luna.Hernandez@soton.ac.uk).

If you have a major difficulty during the course, such as a health problem that prevents you from attending lectures or seriously interferes with your work, you should make sure to discuss this with your Personal Academic Tutor.