

ST236-10 Python for data-analytic tasks

23/24

Department

Statistics

Level

Undergraduate Level 2

Module leader

Matthew Thorpe

Credit value

10

Module duration

10 weeks

Assessment

Multiple

Study location

University of Warwick main campus, Coventry

Description

Introductory description

This module introduces students to the Python programming language, with a particular focus on writing efficient code and its effective management, data-analytic tasks, and mathematical optimisation frameworks in Python.

This module is offered as an optional module to Statistics students and as an unusual option to students from other departments, space permitting.

[Module web page](#)

Module aims

To introduce students to

1. Source-code editors, IDEs and notebooks
2. Effective and collaborative management of code
3. Basic programming concepts and their implementation in Python
4. Data management and data-analytic tasks with Python
5. Visualization in Python

6. Frameworks for mathematical optimization

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.

This module covers the following topics.

1. Basic tools. Source-code editors and IDEs; Installing and working with Python; Version control, repository hosting services and collaboration platforms
2. Data, Data types and Data Structures. Structured, semi-structured, and unstructured data; File formats for data exchange; Data types and structures in Python; Working with Numpy and Pandas; Import/Export of data exchange files in Python
3. Databases. Introduction to relational database systems; Introduction to SQL and SQLite; Basic SQL/SQLite syntax and queries; Creating and manipulating databases in Python; Querying databases in Python
4. Programming concepts. Variables, control flow structures and functions; Variables, mutability and aliasing in Python; Control flow structures in Python; Functions and scope in Python; Exceptions and error handling in Python; Debugging in Python; Classes and programming paradigms; Parallelization
5. Data Wrangling. Introduction to data wrangling; Data wrangling operations in Python; Exploratory data analysis, graphics and data visualization in Python
6. Optimization in Python. Function optimization; linear programming
7. Writing modules and packages. Modules and packages in Python; Documenting code in Python; Test-driven software development

Learning outcomes

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

- Create programs to solve problems.
- Construct readable, valid, reliable and modular code.
- Apply Python programming techniques to manage, store, and visualise data.
- Apply Python programming techniques to data-analytic and/or optimisation tasks.
- Collaborate and disseminate fully documented code with reproducible outputs in the form of Python modules and packages.

Indicative reading list

McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.

Guttag, J. V. (2016). Introduction to Computation and Programming Using Python: With Application to Understanding Data (2nd ed.). MIT Press

Interdisciplinary

This module requires students to develop a balanced facility of rigorous programming and data-analytic skills for solving real-world problems across disciplines.

Subject specific skills

1. Demonstrate facility with data handling and analysis methods in Python
2. Create readable, valid, reliable, and modular code
3. Analyze problems, abstracting their essential information formulating them using appropriate programming concepts to facilitate their solution.
4. Demonstrate programming skills and knowledge of fundamental programming concepts, both explicitly and by applying them to the solution of real-world problems

Transferable skills

1. Problem-solving skills: The module requires students to solve problems and present their conclusions as logical and coherent arguments.
 2. Written communication skills: Students complete written assessments that require precise and unambiguous communication in the manner and style expected in mathematical sciences.
 3. Verbal communication skills: Students are encouraged to discuss and debate formative assessment and lecture material within small-group tutorial sessions. Students can continually discuss specific aspects of the module with the module leader. This is facilitated by statistics staff office hours.
 4. Team working and working efficiently with others: Students are encouraged to discuss and debate formative assessment and lecture material within small-group tutorial sessions.
 5. Professionalism: Students work autonomously by developing and sustaining effective approaches to learning, including time management, organisation, flexibility, creativity, collaboratively and intellectual integrity.
-

Study

Study time

Type	Required
Lectures	20 sessions of 1 hour (20%)
Practical classes	5 sessions of 1 hour (5%)
Private study	35 hours (35%)
Assessment	40 hours (40%)
Total	100 hours

Private study description

Weekly revision of lecture notes and materials, wider reading and practice/programming exercises, working on problem sets and preparing for examination.

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 C

	Weighting	Study time
Group assignment 1	25%	10 hours
A formal group report, to professional standards, presenting the analysis, interpretation and conclusion of the data visualisation task set. All code must be version-controlled, well-documented and reproducible. The target audience are decision-makers who do not necessarily have advanced statistical training. For the purposes of this assessment 500 words is equivalent to one page of text, diagrams, formula or equations. Submitted code will be part of the report's appendix and will not count toward the page limit. This report must not exceed 8 pages in length.		
Group assignment 2	25%	10 hours
A formal group report, to professional standards, presenting the analysis, interpretation and conclusion of the data-analytic and/or optimisation task set. All code must be version-controlled, well-documented and reproducible. The target audience are decision-makers who do not necessarily have advanced statistical training. For the purposes of this assessment 500 words is equivalent to one page of text, diagrams, formula or equations. Submitted code will be part of the report's appendix and will not count toward the page limit. This report must not exceed 8 pages in length.		
Examination	50%	20 hours
You will be required to answer all questions on this examination paper.		

- Answerbook Pink (12 page)
- Students may use a calculator

Assessment group R

	Weighting	Study time
Examination	100%	
You will be required to answer all questions on this examination paper.		

Weighting

Study time

- Answerbook Pink (12 page)

Feedback on assessment

Individual feedback will be provided on problem/programming sheets by class tutors.

Cohort level feedback will be provided for the examination.

Students are actively encouraged to use office hours to build up their understanding and to view all their interactions with lecturers and class tutors as feedback.

[Past exam papers for ST236](#)

Availability

Pre-requisites

To take this module, you must have passed:

- All of
 - [ST117-15 Introduction to Statistical Modelling](#)

Courses

This module is Option list A for:

- USTA-G302 Undergraduate Data Science
 - Year 2 of G302 Data Science
 - Year 2 of G302 Data Science
- USTA-GG14 Undergraduate Mathematics and Statistics (BSc)
 - Year 2 of GG14 Mathematics and Statistics
 - Year 2 of GG14 Mathematics and Statistics
- USTA-Y602 Undergraduate Mathematics, Operational Research, Statistics and Economics
 - Year 2 of Y602 Mathematics, Operational Research, Stats, Economics
 - Year 2 of Y602 Mathematics, Operational Research, Stats, Economics