

MSc in Data Analytics for Government

Course descriptions

January-June 2019

**Faculty of Technology,
Design and Environment**

**School of Engineering,
Computing and Mathematics**

mdatagov@brookes.ac.uk

P08800 Statistics in Government

The aim of this module is to provide a sound overview of the issues and challenges for Official Statistics in the UK.

Learning Outcomes

1. Demonstrate an understanding of the key issues and challenges for the delivery of Official Statistics and apply this knowledge in discussing the relative merits of alternative approaches
2. Critically evaluate the legal framework and international context in which Official Statistics are used

Outline Syllabus

- Overview of Official Statistics
- Political legal and ethical issues
- Codes of Practice
- International requirements and standards
- Classification systems
- The Quality Management Framework
- Building public trust in Official Statistics
- Issues of sharing (legalities) and data confidentiality

Provisional timetable

| Time | Monday | Tuesday | Wednesday | Thursday | Friday |
|-------------|--|--|---|---|-------------------------------|
| 10:00-12:00 | Session 1 Introduction | Session 3 Legal and Administrative Aspects | Session 5 Quality | Session 7 International Standards and Classifications | Session 9 Trust |
| 12:00-13:00 | LUNCH | | | | |
| 13:00-15:00 | Session 2 History of UK Statistics | Session 4 Code of Practice | Session 6 Professional Practice | Session 8 Organisation and Management | Session 10 Big Data |
| 15:00-15:30 | BREAK | | | | |
| 15:30-17:00 | Independent Reading | Independent Reading | Independent Reading | Independent Reading | Independent Reading |

Assessment Tasks

Report

Reading List

| Author | Title | Publisher | Date |
|---|---|---------------|------|
| Fellegi I P | Characteristics of an Effective Statistical System Lecture, Washington | Morris Hansen | 1995 |
| Holt D | Official Statistics, public policy and public trust JRSS A, 171, 2, 323-346 | Wiley | 2008 |
| Laux, Richard, Alldritt, Richard, Dunnell Karen | Evolution of the United Kingdom Statistical System, Statistical Journal of the IAOS 24, pp47 - 59 | IOS Press | 2007 |
| | Statistics and Registration Service Act 2007 | | |

Dates of scheduled courses

10th-14th June 2019 (ONS Newport)

P08802 Survey Fundamentals

The aim of the module is to provide an overview of sampling and estimation fundamentals.

Learning Outcomes

1. Select appropriate methods for sampling from finite populations, including the most common sampling designs.
2. Appropriately estimate finite population parameters and assess estimation errors.
3. In conjunction with others, create effective and non-biased surveys taking account of social and cultural expectations.

Outline Syllabus

- Core statistical survey principles including drawing samples from finite populations and making inferences about population characteristics.
- Understanding different estimation strategies, with consideration of survey errors.
- Probability sampling methods versus non-probability sampling methods.
- Variance estimation, standard errors, and confidence intervals.

Provisional timetable

| Time | Monday | Tuesday | Wednesday | Thursday | Friday |
|-------------|--|--|--|--|---|
| 10:00-12:00 | Session 1 Review of Basic Concepts | Session 3 Selecting a Sample | Session 5 Regression | Session 7 Stratified Random Sampling 2 | Session 9 Systematic Sampling |
| 12:00-12:45 | LUNCH | | | | |
| 12:45-14:45 | Session 2 Surveys and Questionnaires | Session 4 Ratio Estimation | Session 6 Stratified Random Sampling 1 | Session 8 Cluster Sampling | Session 10 Supplementary Topics |
| 14:45-15:00 | BREAK | | | | |
| 15:00-17:00 | Practical Session | Practical Session | Practical Session | Practical Session | Practical Session |

Assessment Tasks

Individual report on group project

Reading List

| Author | Title | Publisher | Date |
|---------------|---|-------------------|------|
| Andres, L. | Designing and Doing Survey Research | Sage | 2012 |
| Fink, A. | How to conduct surveys: a step-by-step guide | Sage | 2016 |
| Fowler, F. J. | Survey Research Methods, 5th Edition | Los Angeles: Sage | 2014 |
| Kent, R. | Data Construction and Data Analysis for Survey Research | Palgrave | 2001 |
| Sapsford, R. | Survey Research, 2nd Edition | Sage | 2011 |

Dates of scheduled courses

20th-24th May 2019 (ONS Newport)

P08820 Introduction to Machine Learning

The module aims to provide the students with the principles of computer learning and its applications. It covers the fundamentals of machine learning methodologies, implementations and analysis methods appropriate for machine learning applications. The module will enable students to analyse a machine learning problem, critically evaluate the different approaches that are available, and create an effective solution.

Learning Outcomes

1. Evaluate and articulate the issues and challenges in machine learning, including model selection, complexity and feature selection.
2. Demonstrate a working knowledge of the variety of mathematical techniques normally adopted for machine learning problems, and of their application to creating effective solutions.
3. Critically evaluate the performance and drawbacks of a proposed solution to a machine learning problem.
4. Create solutions to machine learning problems using appropriate software.

Outline Syllabus

- Foundations of machine learning and relation to Artificial Intelligence (AI)
- Supervised learning
- Unsupervised learning
- Artificial Neural Networks

Provisional timetable

| Time | Monday | Tuesday | Wednesday | Thursday | Friday |
|-------------|---|-------------------------------------|--|-------------------------------------|---|
| 10:00-12:00 | Session 1 Foundations of Machine Learning | Session 3 Clustering | Session 5 K-nearest neighbour | Session 7 Perceptron | Session 9 AI Case Study 2 |
| 12:00-12:45 | LUNCH | | | | |
| 12:45-14:45 | Session 2 Supervised versus unsupervised learning | Session 4 ML Case Study 1 | Session 6 Decision trees and forests | Session 8 Neural networks | Session 10 Supplementary Topics |
| 14:45-15:00 | Practical Session | | | | |
| 15:00-17:00 | Practical Session | Practical Session | Practical Session | Practical Session | Practical Session |

Assessment Tasks

Report on machine learning problem

Reading List

| Author | Title | Publisher | Date |
|-----------------|---|-------------|------|
| Bishop, C. M. | Pattern Recognition and Machine Learning | Springer | 2006 |
| Flach, P. A. | Machine Learning: The Art and Science of Algorithms That Make Sense of Data | CUP | 2012 |
| Harrington, P. | Machine Learning in Action | Manning | 2012 |
| Mitchell, T. M. | Machine Learning | McGraw Hill | 1997 |

Dates of scheduled courses

21st-25th January 2019 (ONS Newport)

24th-28th June 2019 (ONS Titchfield)

P08822 Introduction to Distributed Systems

This module will provide an overview of processing data at large scale and parallel processing. It will introduce Hadoop and Spark and the use of parallel processing paradigms. It will also extend to Data Analytics (distributed ETL and machine learning e.g. more advanced Hadoop & Spark) including:

- Data querying using remote database management system (RDMS)/NoSQL databases
- data processing pipeline
- database design using remote database management system (RDMS)/NoSQL databases
- cloud computing

Learning Outcomes

1. Design and implement a distributed system application utilising modern design paradigms and technologies
2. Critically evaluate distributed systems performance characteristics on complex data
3. Utilise distributed systems concepts to enhance the performance of related activities

Outline Syllabus

- Distributed Systems Theory
- Database management system (DBMS) theory and practice
- Hadoop
- Spark
- Cloud Computing

Provisional timetable

| Time | Monday | Tuesday | Wednesday | Thursday | Friday |
|-------------|--|---|--|--|---|
| 10:00-12:00 | Basic concepts of distributed computing | Infrastructures for large scale distributed computing (Part 2: Distributed and NoSQL databases) | Distributed parallel programming models (Part 1: MapReduce) | Service oriented architectures: Big and RESTful Web Services | Cloud computing paradigms: (Part 2: SaaS, Monolithic and Microservices architectures, container technologies) |
| 12:00-13:00 | LUNCH | | | | |
| 13:00-14:45 | Infrastructures for large scale distributed computing (Part 1: Hadoop distributed file system and related tools) | Data models in distributed applications: Structured, semi-structured and Unstructured) | Distributed parallel programming models (Part 2: Actor, Reactive, and Agent) | Cloud computing paradigms: (Part 1: IaaS, PaaS and virtualization) | Case study |
| 14:45-15:00 | BREAK | | | | |
| 15:00-16:30 | Independent Study and Reading | Independent Study and Reading | Independent Study and Reading | Independent Study and Reading | Independent Study and Reading |

Assessment Tasks

Analyse a complex distributed system with respect to performance and reliability to produce recommendations on its use in a given scenario (40%)

Design and implement a distributed system to a given specification (60%)

Reading List

| Author | Title | Publisher | Date |
|---|--|------------------|-------------|
| Sadalage, P. J. and Fowler, M. | NoSQL Distilled | Addison-Wesley | 2013 |
| Eadline, D., Stella, C., Mendelevitch, O. | Practical Data Science with Hadoop® and Spark: Designing and Building Effective Analytics at Scale | Addison-Wesley | 2016 |
| Harrison, G. | Next Generation Databases: NoSQL, NewSQL, and Big Data | Apress | 2016 |
| Kleppmann, M. | Designing Data-Intensive Applications, 1st Edition | O'Reilly Media | 2017 |
| Varela, C. A. | Programming Distributed Computing Systems: A Foundational Approach | MIT Press | 2017 |

Dates of scheduled courses

20th-24th May 2019 (ONS Titchfield)

P08825 Data Visualisation

This module will build on the basic data visualisations introduced in the compulsory modules. It will cover information design, interaction design and user engagement; state of the art tools to build useful visualisations for different types of data sets and application scenarios; mapping.

Learning Outcomes

1. Critically analyse data visualisation approaches with respect to human sensory modalities.
2. Create appropriate visualisations for temporal, dynamic, and high dimensionality data.
3. Devise methodologies for data interaction to facilitate exploratory data analysis.

Outline Syllabus

- The human perceptual system
- Representing complex data
- Representing dynamic and temporal data
- Exploratory data analysis
- Visualisation software tools

Provisional timetable

| Time | Monday | Tuesday | Wednesday | Thursday | Friday |
|-------------|-----------------------------------|--|-------------------------------|--------------------------------------|---------------|
| 10:00-12:00 | Introducing Data Visualisation | Analysing the purpose of the visualisation | Data Visualisation Principles | Spatial data, networks | Temporal Data |
| 12:00-13:00 | LUNCH | | | | |
| 13:00-14:45 | Processing Data for visualisation | Software to process Data | Visualisation tables | Visualisation software tools | Case study 2 |
| 14:45-15:00 | BREAK | | | | |
| 15:00-16:30 | Programming exercises | Programming: Exercises with data frames | A case study | Visualisation with case study | |

Assessment Tasks

Part 1: Evaluate a range of different data visualisations on a given set of multi variate, dynamic, data

Part 2: Develop a report to visualise data in different formats

Reading List

| Author | Title | Publisher | Date |
|--|---|-------------------|------|
| Atmajitsinh Gohil | R Data Visualization Cookbook | Packt Publishing | 2015 |
| Tamar Muzner | Visualization Analysis and Design | CRC Press | 2015 |
| Andy Kirk | Data Visualisation | Sage Publications | 2016 |
| Mathew Ward, George Grinstein, Daniel Keim | Interactive Data Visualization, 2nd Edition | CRC Press | 2015 |

Dates of scheduled courses

21st-25th January 2019 (ONS Titchfield)