MSIN0010 Data Analytics I
Term 1 2020

Lecturer:	Dr Adam Smith
Email:	a.smith@ucl.ac.uk
Office Hours:	Friday 1:00 - 2:00 pm

TA:	Mr Xitong (Tony) Liang
Email:	xitong.liang.18@ucl.ac.uk
Office Hours:	Monday 11:30 am – 1:00 pm

Module Description

Data Analytics I introduces students to how organizations use data and analytics to create value and improve performance, trains them to use selected statistical data analytics and data mining tools, and introduces them to elements of the statistical theory and algorithms that underpin those tools.

The context for the module is management in complex, innovation-intensive, data-driven environments. The explosion in the volume and range of internal and external data available to managers and the development of new data analytics tools is having a major impact on how people identify, formulate, and solve management problems.

During the module, students will manipulate example data sets and use basic data collection tools and APIs to source data from publicly available data sources.

Learning Outcomes

Upon successful completion of the module, students will be able to:

- Understand how organizations use data and analytics to create value and improve performance.
- Understand and apply founding probability and statistical theory to data analysis.
- Understand and apply information theory and data mining theory to data classification and data clustering problems.
- Characterize and critically assess the quality of data sets and their limitations in the context of data-driven decision-making.
- Use selected tools (Excel and R) to analyze and visualize data.
- Understand key elements of the theory, technology, and algorithms that underpin the tools used.

Assessment

1. (Term 1) DataCamp Assignments – 20%
2. (Term 2) Group Project – 20%
3. (Term 3) Exam – 60%

Required Materials

1. Online Textbook: https://www.adamnsmith.com/MSIN0010/
2. Data Camp: https://www.datacamp.com/
3. RStudio: https://rstudio.cloud
4. All lecture notes and additional readings will be posted on Moodle

Additional Resources

Textbooks on Probability and Inference

- *All of Statistics* (2004) by L. Wasserman
- *Introduction to Probability* (2019) by J. Blitzstein and J. Hwang [http://probabilitybook.net/]

Textbooks on Regression and Machine Learning

- *An Introduction to Statistical Learning* (2013) by G. James, D. Witten, T. Hastie, and R. Tibshirani [http://faculty.marshall.usc.edu/gareth-james/ISL/]
- *Business Data Science* (2019) by M. Taddy

Popular Books on Statistics, Econometrics, and Measurement in the Social Sciences

- *The Art of Statistics* (2019) by D. Spiegelhalter
- *The Book of Why* (2018) by J. Pearl and D. Mackenzie
- *Freakonomics* (2005) by S. Levitt and S. Dubner
- *Mastering 'Metrics* (2014) by J. Angrist and J. Pischke
- *The Signal and the Noise* (2012) by N. Silver

Schedule

UNIT I: DATA

1. October 5-9
   - Videos: **Introduction to Data Analytics** (Brief History of Data Analytics, Data Sets)
   - Reading: *The Advantages of Data-Driven Decision-Making, Big Data - The Management Revolution*
   - DataCamp: Spreadsheet Basics (chapter 1), Introduction to R (chapter 1)
2. October 12-16
   - Videos: **Summarizing Data** (Data Visualization, Summary Statistics)

- Reading: *How R Helps Airbnb Make the Most of its Data*
- DataCamp: Introduction to the Tidyverse (4 chapters), Introduction Data Visualization with ggplot2 (4 chapters)

UNIT II: PROBABILITY

3. October 19-23
   - Videos: **Probability** (Probability Theory, Random Variables, Probability Distributions)
   - Reading: *Probability - The Language of Uncertainty*
4. October 26-30
   - Videos: **Probability** (Expectations, Central Limit Theorem)

UNIT III: STATISTICAL INFERENCE

5. November 2-6
   - Videos: **Estimation** (Point Estimation, Confidence Intervals)
   - DataCamp: Foundations of Inference (4 chapters)
6. November 9-13 (READING WEEK)
7. November 16-20
   - Videos: **Testing** (Hypothesis Testing)

UNIT IV: STATISTICAL MODELS AND MACHINE LEARNING

8. November 23-27
   - Videos: **Regression** (Linear Regression, Regression Trees)
   - DataCamp: Supervised Learning in R - Regression (chapters 1 and 5)
9. November 30 - December 4
   - Videos: **Classification** (NN Algorithms, Logistic Regression, Classification Trees)
   - DataCamp: Supervised Learning in R - Classification (chapters 1 and 4)
10. December 7-11
    - Videos: **Model Selection**
    - Reading: *What Do We Do About the Biases in AI?*
    - DataCamp: Supervised Learning in R - Regression (chapter 2)
11. December 14-18
    - Videos: **Clustering**
    - DataCamp: Unsupervised Learning in R (chapter 1)