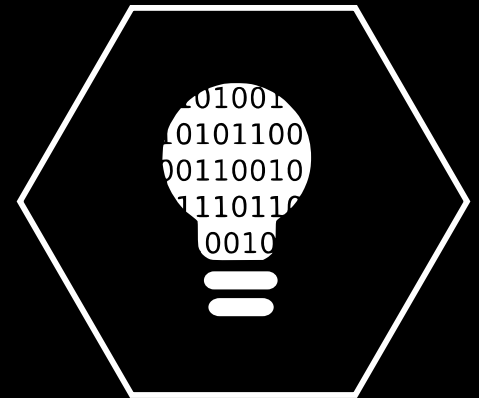# Data Science Foundations
Course information sheet 2022-23

Full course, 10 weeks

This course introduces students to data analytics and data science as well as different approaches to learning from data and provides an introduction to statistical model-based inference.

Prerequisite Knowledge

Learners should have a basic understanding of mathematics including matrix algebra and calculus, for example differentiation. Learners should also have basic experience with the R programming language (e.g. data management and plotting).

Intended Learning Outcomes

By the end of this course learners will be able to:

- explain different types of data and data structures and discuss advantages and challenges of using data of different types in a given context;
- describe different ways of collecting data and discuss advantages and challenges of using data obtained from different sources in a given context;
- describe and visualise structured and unstructured data of different types using suitable summaries and plots;
- explain different approaches to learning from data and discuss their advantages and disadvantages in a given context;
- define and contrast *population* and *sample*, *parameter* and *estimate*;

- implement these statistical methods using the R computer package.
- write down and justify criteria required of 'good' point estimators, and check whether or not a proposed estimator within a stated statistical model satisfies these criteria;
- apply the principle of maximum likelihood to obtain point and interval estimates of parameters in statistical models, making appropriate use of numerical methods for optimisation;
- formulate and carry out hypothesis tests in Normal models, as well as general likelihood-based models, correctly using the terms *null hypothesis*, *alternative hypothesis*, *test statistic*, *rejection region*, *significance level*, *power*, *p-value*;

Syllabus

Week 1
- What is learning from data?
- Data sources, structures and data types
- Collecting and collating data

Week 2
- Summarising and visualising data
- Quality assuring data
- Exploring relationships in data

Week 3
- What is statistical inference?
- A framework for hypothesis testing
- Interpreting confidence intervals and p-values

Week 4
- Calculating confidence intervals and constructing hypothesis tests for one/two sample problems
- Computing the intervals and tests in R
- Interpreting output from confidence intervals and hypothesis tests

Week 5 (sample material)
- Properties of point estimators
- The idea and concept of maximum likelihood
- Maximum likelihood estimation for discrete distributions

Mid-term week break

Week 6
- Maximum likelihood for continuous distributions
- Maximum likelihood estimation on a boundary
- Numerical optimisation
- Properties of point estimators

Week 7
- Definitions of relative likelihood and relative log-likelihood
- Likelihood intervals
- Large sample properties to obtain confidence intervals
- Interpreting the results of these intervals

Week 8
- Maximum likelihood for the normal distribution and multiple independent populations
- The Hessian matrix
- Properties of Maximum Likelihood Estimators

Week 9
- Approximate confidence intervals to compare parameters from independent populations
- Interpreting the results of these intervals
- Comparing hypotheses using likelihood
- Type I and Type II errors and statistical power

Week 10
- Large sample properties for a Generalised Likelihood Ratio Test (GLRT)
- Applying and interpreting results from a GLRT
- Deriving a GLRT for the multinomial distribution

Supplementary Material
- Motivation for Bayesian inference
- Using Bayes' theorem to obtain posterior distributions
- Visualising prior and posterior distributions and likelihoods in R

"Masterclass in how to deliver a teaching module. Course notes were clear and concise with tasks that were relevant and required application of knowledge. Videos always clearly explained."

Online Learning
- Weekly live sessions with tutor(s)
- Weekly learning material (reading material, videos, exercises with model answers)
- Bookable one-to-one sessions with tutor(s)

Textbooks

Panik, M (2012) Statistical inference: a short course

Lee, H (2014) Foundations of applied statistical methods

Held, L & Bové, D (2014) Applied statistical inference: likelihood and Bayes

Gergely, D (2015) Mastering data analysis with R: gain clear insights into your data and solve real-world problems.

Assessment
(for credit only)

This will typically be made up of 5 pieces of assessment, including online quizzes, an individual project and an online test.

DATA ANALYTICS
GLASGOW

School of Mathematics and Statistics

University of Glasgow

http://gla.ac.uk/mdatagov

http://gla.ai

Email:

maths-stats-analyticscpd@glasgow.ac.uk

Software
To take our courses please use an up-to-date version of a standard browser (such as Google Chrome, Firefox, Safari, Internet Explorer or Microsoft Edge) and a PDF reader (such as Acrobat Reader). Learning material will be distributed through Moodle. We encourage all learners to install R and RStudio and we provide detailed installation instructions, but learners can also use free cloud-based services (RStudio Cloud). Learners need to install Zoom for participating in video conferencing sessions. We recommend the use of a head set for video conferencing sessions.