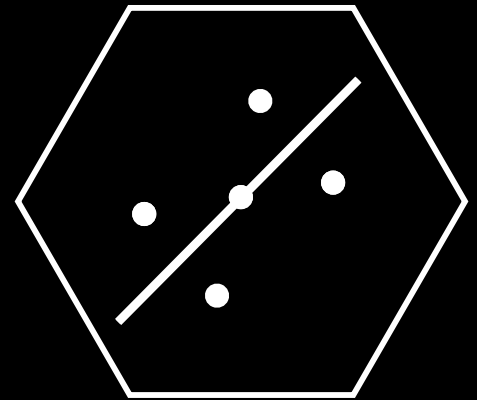


# Predictive Modelling

## Course information sheet 2022-23

Full course, 10 weeks



This course introduces predictive modelling using multiple linear regression. It presents the theory underpinning the normal linear model accommodating continuous and categorical predictors and non-linear effects. In addition it contrasts common methods for model assessment and selection.

### Prerequisite Knowledge

Learners should have a basic understanding of mathematics including matrix algebra and calculus, for example differentiation. Learners should also have basic experi-

ence 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:

- formulate normal linear models in vector-matrix notation and apply general results to derive ordinary least squares estimators;
- construct a design matrix incorporating categorical covariates or covariates with a nonlinear effect;
- derive, evaluate and interpret point and interval estimates of model parameters;
- derive, evaluate and interpret hypothesis tests and confidence and prediction intervals for the response at particular values of the explanatory variables;
- assess the assumptions of a normal linear model;
- implement these statistical methods using R.
- fit multiple linear regression models;
- make use of and critique different methods for assessing the performance of a predictive model and use these for model or variable selection;
- implement fitting and model selection of multiple linear regression methods using R;
- identify scenarios where data may be considered to be smooth functions and apply suitable non-linear techniques.

# Syllabus

## Week 1

- Understand the scope of predictive modelling
- Definition of a statistical model
- Definition of a linear regression model

## Week 2 (sample material)

- Fitting a simple normal linear model from first principles
- Least squares estimation in vector matrix notation

## Week 3

- Defining models with continuous and categorical variables in vector matrix notation
- Interpreting regression coefficient estimates
- Fit a normal linear regression model in R

## Week 4

- Define residuals of the normal linear model
- State the assumptions of the normal linear model
- Assess model assumptions using residual plots

## Week 5

- Calculate confidence and prediction intervals for a specified confidence level
- Perform hypothesis testing to test the significance of coefficients in a normal linear model
- Describe the analysis of variance

table for a normal linear model

## Mid-term week break

## Week 6

- Assessing the significance of an interaction
- Perform variables selection with several explanatory variables

## Week 7

- Hierarchical Models
- Forward Selection and Stepwise Regression
- Criterion-based procedures for model selection

## Week 8

- Demonstration of model selection in R with several examples

## Week 9

- Describe non-parametric regression modelling
- State, explain and compare several methods of smoothing
- Construct simple splines

## Week 10

- Write down simple spline models in vector matrix notation
- Assess fitting criteria for regression splines and penalised regression spline.
- Specifying an additive model and describing methods used to fit an additive model
- Using existing R libraries to fit additive models

*“Challenging and topics covered are extensive but relevant for developing skills in data analytics.”*

## Online Learning

- Weekly live sessions with tutor(s)
- Weekly learning material (reading material, videos, exercises with model answers)
- One-to-one sessions with tutor(s) on request

## Textbooks

Faraway, J J (2015) Linear models with R, 2nd ed, CRC Press.  
Chatterjee, S and Hadi A S (2012) Regression analysis by example, 5th ed, John Wiley & Sons, Inc.  
Ramsay, J O and Silverman, B W (2005) Functional Data Analysis, 2nd ed, Springer.

## Assessment

(for credit only)

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

## 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.

DATA ANALYTICS   
**GLASGOW**

School of Mathematics and Statistics

University of Glasgow

<http://gla.ac.uk/>

[mdatagov](http://mdatagov)

<http://gla.ai>

Email:

[maths-stats-analyticscpd@glasgow.ac.uk](mailto:maths-stats-analyticscpd@glasgow.ac.uk)