University of Glasgow

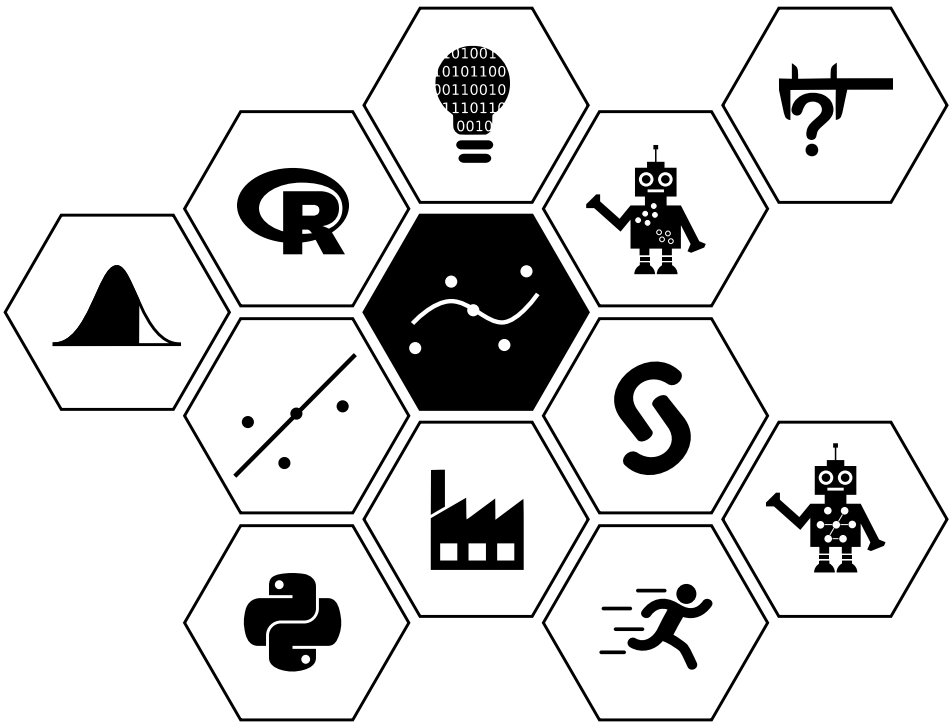
Advanced Predictive Models

Tereza Neocleous

Academic Year 2020-21

Week 1:

Introduction to Advanced Predictive Models



DATA ANALYTICS GLASGOW

General Information: Advanced Predictive Models

Course description

This course is about models which can account for a non-normal distribution of the response and/or the fact that data is not independent, but correlated.

The course lecturer is Tereza Neocleous (email: tereza.neocleous@glasgow.ac.uk).

Introduction to Advanced Predictive Models

https://youtu.be/ApZu75PkEXw

Duration: 1m36s

Aims

The aims of this course are:

- to provide an overview of different generalisations of linear regression models
- to acquaint students with the theory of exponential families
- to introduce generalised linear models
- to introduce the concept of a time series and to present a range of approaches for representing trends and seasonality
- to illustrate how temporal correlation can be incorporated into a regression model
- to illustrate how random effects can be incorporated into a regression model

Intended learning outcomes

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

- explain and derive key aspects of the theory of exponential families and generalised linear models;
- make correct use of models with various link functions and link distributions such as models for discrete data;
- determine whether a time series exhibits any evidence of a trend, seasonality or short-term correlation;
- define the class of ARIMA probability models;
- determine an appropriate model for a data set from the class of ARIMA models;
- predict future values for a given time series;
- make correct use of regression models assuming correlated residuals as well as models based on generalised estimation equations;
- explain the notion of a random effect, why and when it is useful and, in particular, how it differs from a fixed effect;
- make correct use of hierarchical models with random effects.

For more detailed information please see the course catalogue entry for STATS 5073.

Course materials and schedule

Course materials will be released on the course Moodle page and on Microsoft Teams. The course is split into Part 1 and Part 2 in Teams. Please use the team code **84vroua** to join Part 1 and **m2u9cgm** to join Part 2.

Please note that while the notes will be available from both Moodle and Teams, assessments will be submitted on Moodle.

The course is split into two parts, Part 1 on Generalised Linear Models, and Part 2 on time series analysis and models for correlated responses. There is a midterm break between the two parts during the week of 7th June 2021.The course will consist of the following topics:

Part 1:

- Week 1 (beginning 3 May 2021) - Introduction and course information
- Week 2 (beginning 10 May 2021) - Generalized linear models: what are they, estimation and inference
- Week 3 (beginning 17 May 2021) - Models for binary/binomial response
- Week 4 (beginning 24 May 2021) - Models for ordinal/nominal response
- Week 5 (beginning 31 May 2021) - Models for counts

Part 2:

- Week 6 (beginning 14 June 2021) - Introduction to time series data, trend and seasonality
- Week 7 (beginning 21 June 2021) - Autoregressive and moving average processes and the ARIMA approach
- Week 8 (beginning 28 June 2021) - Forecasting in time series analysis
- Week 9 (beginning 5 July 2021) - Models for correlated observations: random effects, mixed model approach
- Week 10 (beginning 12 July 2021) - Models for correlated discrete responses

Assessment and feedback calendar

Assessment is 100% coursework as shown below. If you are having issues with any of the deadlines, please email Tereza to request an extension.

Assessment Component	Grade %	Released	Due	Feedback
Quiz 1	20%	24 May 2021	6 June 2021	Upon quiz submission/quiz closed
Assignment 1	25%	24 May 2021	17 June 2021	5 July 2021
Quiz 2	20%	28 June 2021	11 July 2021	Upon quiz submission/quiz closed
Assignment 2	25%	28 June 2021	13 July 2021	30 July 2021
Quiz 3	10%	12 July 2021	25 July 2021	Upon quiz submission/quiz closed

Interactive Zoom sessions

There will be live Question & Answer sessions on Tuesdays from 13:00-14:00 UK time (BST) (invite link here). These are an opportunity to ask any questions you may have about the course. They will be recorded with the recordings made available later for anyone who cannot attend.

Q&A chat: I would also encourage you to use the Teams chat functionality to post and answer questions. This can often get your questions answered faster than waiting for one of the scheduled Zoom sessions.

Resources

This course will use the following resources which will be made available on the course Moodle and Teams pages:

- A probability distribution sheet
- Statistical tables

The course will be self-contained in the learning material. However, for each week we will point to chapters in the following books that you might wish to consult for additional material:

- Faraway - **Extending linear models with R: generalized linear, mixed effects and nonparametric regression models**
- Zuur et al. - **Mixed effects models and extensions in ecology with R**
- Fahrmeir et al. - **Regression: models, methods and applications**
- Ruppert et al. - **Semiparametric Regression**
- Venables & Ripley - **Modern Applied Statistics with S**
- Hyndman & Athanasopoulos - **Forecasting: Principles and Practice**