1.2 Scope of the course

Scope of the course

This slide gives you an overview of the flavour of this course, the list of topics covered, models, estimations techniques, etc. When diving in the lessons you will find plenty of real-world applications of the methods covered.

Week 1:

Introduction to regression modelling and key statistical concepts that we will use throughout the course. This includes:

- estimation methods,
- maximum likelihood estimation,
- least squares estimation,
- exponential family of distributions,
- prediction,
- inference.

Week 2:

We dive in the world of regression modelling! The focus is on the popular linear models. This includes:

- Simple linear regression
- Multiple linear regression
- Estimation and hypothesis testing
- Confidence intervals and prediction intervals
- ANOVA (Analysis of Variance) used for a continuous response variable and categorical explanatory variables (factors)
- ANCOVA (Analysis of Covariance) is used when at least one of the explanatory variables is continuous
- A generalisation

Week 3:

We move towards a more flexible regression model: the Generalised Linear Model or GLM. This allows to tailor models to specific problems and types of data. This includes:

- Logistic regression (special case)
- Poisson regression (special case)

- log-linear regression (special case)
- Methods of estimation
- Model fitting
- Statistical inference

Week 4:

We look into how to select the best model amongst a pool of models and estimate the error that comes with the selected model based on new (test) data. This includes:

- Resampling methods: how to estimate the estimation error, cross validation
- Variable selection: find the best subset of predictor for a single model.
- Shrinkage Methods: by shrinking the coefficient estimates towards zero, their variance is being reduced. Ridge regression and Lasso are two best-known examples of shrinkage methods;

Week 5:

We learn about nonlinear Regression! Nonlinear models are more complex in terms of interpretation and inference. However, their advantage lies in their predictive power. This includes:

- Polynomial regression: extends the linear model by raising the original predictors to a power. For instance, the cubic regression will have three variables: X, X^2 and X^3 as predictors.
- Step functions: cuts the range of a variable into K distinct regions, which produces a piecewise constant fit.
- Regression splines: extension of polynomials and step functions and involve dividing the range of X into K distinct regions. Within each region a polynomial function is fit to data.
- Smoothing splines: similar to regression splines and result from minimizing the residual sum of squares subject to smoothing penalty.
- Local regression: the regions are allowed to overlap.

Week 6:

Last week of learning! *We discover Generalised Additive Models (GAMs).* This includes:

- Additive models.
- Generalisation of additive models. We have two or more explanatory variables for which we use smooth functions to explain the response variable via the regression model.