Semester 1 Statistics Short courses

Course: STAA0001 - Basic Statistics

Blackboard Site: STAA0001

Dates: Sat. March 12th and Sat. April 30th (9 am – 5 pm)

Room ATC325

Assumed Knowledge: None

Software used: SPSS

Maximum 20 students

Course Description

Statistical techniques as listed below will be covered with an emphasis on the interpretation and reporting of these results.

Day 1: Exploratory Data Analysis Sat March 12th (9 am – 5 pm)
- Levels of measurement of data
- Graphical Analysis: bar charts, pie charts, boxplots, histogram, stem and leaf, scatterplot, clustered and stacked barcharts
- Descriptive Statistics: mode, mean, median, standard deviation, range, IQR, Pearson’s r

Day 2: Introduction to Inference Sat May 1st (9 am – 5 pm)
Introduction to the basic concepts of inference: confidence interval, significance, p-values, effect size statistics
- Sampling distribution of the mean
- z-test and confidence interval for the mean when the population standard deviation is known
- t-tests: one sample, paired and independent
- chi-square

Online quizzes will be available for self-assessment.
Course: STAA0002 - Simple Linear Regression and ANOVA

Blackboard Site: STAA0002

Dates: Sat 19th March and Sat 16th April (9 am – 5 pm)

Room ATC325

Pre-requisites: Basic Statistics (STAA0001)

Software used: SPSS

Maximum 20 students

Course Description:

Statistical techniques as listed below will be covered with an emphasis on the interpretation and reporting of these results.

Day 1: Simple Linear Regression Sat 19 March (9 am – 5 pm)
- Correlation: Pearson, Spearman and Kendall’s tau-b
- Power Analysis for Pearson correlation
- Simple linear regression analysis
- Assumptions and inference for regression
- Common pitfalls of regression
- Data transformations

Day 2: ANOVA Sat 16 April (9 am – 5 pm)
- One way analysis of variance (ANOVA)
- Repeated measures analysis of variance
- Factorial analysis of variance
- Reporting of ANOVA results
- Power analysis and effect size statistics such as eta squared and omega squared

*Online quizzes will be available for self-assessment.*
Course: STAA0003A - Intro to SPSS

Blackboard Site: STAA0003

Dates: Sun 6/3 (9 am – 5 pm)

Room ATC325

Assumed Knowledge: None

Software used: SPSS

Maximum 20 students

Course Description:

On completion of this course, students should be able to use the menus in the data analysis package IBM SPSS Statistics to take data such as that obtained from questionnaires and administrative records or from existing electronic formats and establish appropriate computer files from which basic statistical summaries, graphs and reports can be produced. It will also show the importance of integrating the development of your data collection instrument, such as a questionnaire, with your computer program.

Topics covered will include

- Introduction to IBM SPSS Statistics
- IBM SPSS Statistics data definition
- Establishing an SPSS data file from a questionnaire.
- Basic data analysis in SPSS
- Computing new variables in SPSS
- Recoding, selecting data in SPSS
- Graphing in SPSS

*Online quizzes will be available for self-assessment.*
Course: STAA0003B - Further SPSS

Blackboard Site: STAA0003

Dates: Sun 10/4 (9 am – 5 pm)

Room  ATC325

Assumed Knowledge: Intro to SPSS

Software used: SPSS

Maximum 20 students

Course Description:

On completion of this course, students will become more efficient in their use of SPSS and expand their knowledge of SPSS data handling procedures.

Topics covered will be chosen from:
- An introduction to SPSS syntax
- Computing, recoding and selecting data in SPSS using syntax.
- Dates in SPSS
- Merging files
- Managing complex data files
- SPSS Tables

*Online quizzes will be available for self-assessment.*
Course: STAA0004 - Survey Design

Blackboard Site: STAA0004

Dates: Sun 20 March (9 am – 5 pm)

Room: ATC325

Assumed Knowledge: Basic Statistics Day 1

Software used: None

Course Description:

You will acquire skills and knowledge in the collection of survey, observational, experimental and secondary data; developing a questionnaire, and writing of descriptive reports.

Topics will include:
- Introduction to survey research
- The basics of survey sampling
- How to collect survey data
- Making the most of secondary data
- Developing a questionnaire
- Introduction to scale development
- Coding and cleaning survey data

Online quizzes will be available for self-assessment.
Course: STAA0005A - Multiple Linear Regression  (9 hrs)

Blackboard Site: STAA0005

Dates: Mondays 29/2, 7/3, 14/3 (5.30 pm – 8.30 pm)

Room  EN409

Assumed Knowledge: ANOVA and Simple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description:

In Multiple Regression you will look at simple linear regression and multiple regression using three different strategies (standard regression, stepwise regression and hierarchical regression). Particular attention is paid to report writing, assumption checking, outlier checking and tests for mediation. Make sure that you have access to SPSS and please revise the relevant material for the simple linear regression and ANOVA short course beforehand.

Online quizzes will be available for self-assessment.
Course: STAA0005B - Factor Analysis  (9 hrs)

Blackboard Site: STAA0005

Dates: Mondays 21/3, 4/4, 11/4 (5.30 pm – 8.30 pm)

Room  EN409

Assumed Knowledge: ANOVA and Simple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description:

Factor Analysis covers exploratory factor analysis (EFA). The various methods for extracting and rotating factors are discussed as are the interpretation of factors and the creation of factor scores and summated. EFA is a descriptive technique. That is, it is designed to help us understand and explain patterns in the data, without making any formal predictions about what results will look like. However, it is not our data’s job to tell us what its underlying structure is and a sound factor analytic study will begin with a great deal of prior thinking about the nature of the concept that we want to understand, appropriate indicators of that concept, appropriate population, and how results of factor analysis will be used. So even before we begin data collection, let alone data analysis, we will have an expectation about what the results might look like. The job of the data is then to show us how well our expectations are reflected in the ‘real world’. The results of exploratory factory analysis can then be used inform future hypotheses. These hypotheses are subsequently tested using confirmatory factor analysis (CFA), which is conducted within the structural equation modelling framework (not covered in this subject).

Make sure that you have access to SPSS and please revise the relevant material for the ANOVA and Simple Linear Regression short course beforehand.

*Online quizzes will be available for self-assessment.*
Course: STAA0005C - MANOVA (9 hrs)

Blackboard Site: STAA0005

Dates: Mondays 18/4, 2/5 (5.30 pm – 8.30 pm)

Room EN409

Assumed Knowledge: ANOVA and Simple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description:

MANOVA examines between subjects, within subjects and mixed multivariate analysis of variance. Particular attention is paid to assumption checking, the testing of specific contrasts and report writing. Make sure that you have access to SPSS and please revise the relevant material for the ANOVA and Simple Linear Regression short course beforehand.

*Online quizzes will be available for self-assessment.*
Course: STAA0005D - Binary Logistic Regression

Blackboard Site: STAA0005

Dates: Monday 9/5 (5.30 pm – 8.30 pm)

Room EN409

Assumed Knowledge: Multiple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description:

Logistic Regression is a specialized form of regression that is designed to predict and explain a binary (two-group) categorical variable rather than a metric dependent measure. Its variate is similar to regular regression and made up of metric independent variables. It is less affected than discriminant analysis when the basic assumptions, particularly normality of the independent variables, are not met. The purpose of this course is to provide a guide how to analyse data, interpreting SPSS output for logistic regression analysis, with emphasis on the results that are of essence to report writing. Sample report for logistic regression analysis will be provided.

Make sure that you have access to SPSS and please revise the relevant material for the ANOVA and Simple Linear Regression short course beforehand.

Online quizzes will be available for self-assessment.
Course: STAA0005E - Discriminant Analysis  (3 hrs)

Blackboard Site: STAA0005

Dates: Monday 16/5 (5.30 pm – 8.30 pm)

Room EN409

Assumed Knowledge: ANOVA and Simple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description:

Discriminant function analysis (DFA) belongs to a family of classification techniques, which are used to predict group membership on a categorical dependent variable (DV, also referred to as grouping variable) from a combination of continuous independent variables (IV, also referred to as predictors). Classification techniques can be divided into those that use linear combinations of predictors to distinguish between groups and those that use non-linear transformations of the combined predictors to distinguish between DV groupings. DFA is an example of linear classification, where group membership on a categorical DV is predicted from a linear combination of continuous predictors. Logistic regression and probit regression are examples of non-linear classification. In this course we will take a closer look at non-linear classification in the context of logistic regression. Related to classification analysis is clustering, which is used to identify which conceptually meaningful groups (if any) exist in the data.

Make sure that you have access to SPSS and please revise the relevant material for the ANOVA and Simple Linear Regression short course beforehand.

*Online quizzes will be available for self-assessment.*
Course: STAA0010A - Multiple Regression Extensions and the General Linear Model

Blackboard Site: STAA0010

Dates: Wed 2/3, 9/3, 16/3 (5:30pm – 8:30pm)

Room ATC325

Assumed Knowledge: Multiple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description

Statistical techniques as listed below will be covered with an emphasis on the interpretation and reporting of these results.

A review of multiple linear regression with special attention to assumptions, unusual point identification and multicollinearity. Different regression techniques are introduced and tests for mediation and moderation are illustrated. A variety of methods for improving the fit of regression models are provided. Methods for weighted regression, nonlinear regression methods and the General Linear Model are then introduced, always assuming that residuals are independent and normally distributed.

Online quizzes will be available for self-assessment.
Course: STAA0010B - Generalised Linear Models and Loglinear Analysis

Blackboard Site: STAA0010

Dates: Wed 23/3, 30/3 (5:30pm – 8:30pm)

Room: ATC325

Assumed Knowledge: Multiple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description

Statistical techniques as listed below will be covered with an emphasis on the interpretation and reporting of these results.

When normal assumptions are no longer valid the Generalised Linear Model is used. These models are introduced and then we look particularly at categorical variables. Starting with Crosstab analyses we learn how to define residuals that help us to interpret relationships between categorical variables. Special measures of association are developed for particular types of categorical variables. Finally multi-order crosstab tables are introduced together with the loglinear analyses required to test for multi-way interaction effects.

*Online quizzes will be available for self-assessment.*
Course: STAA0010C - Advanced Logistic Regression

Blackboard Site: STAA0010

Dates: Wed 6/4, 13/4 (5:30pm – 8:30pm)

Room ATC325

Assumed Knowledge: Multiple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description

Statistical techniques as listed below will be covered with an emphasis on the interpretation and reporting of these results.

Binary logistic regression is a special generalised linear model for binary response variables which uses a logistic link function. We learn how to interpret odds and odds ratios and then show how binary logistic regression is used in practice to fit models with more than one predictor variable. Univariate binary logistic regression models are first fitted using each predictor in turn with a multiple binary logistic regression model to follow. This allows us to test for mediation. Finally, ROC curves and the Hosmer-Lemeshow test are used to assess goodness of fit.

Ordinal logistic with an ordinal response variable are then introduced and tested for “parallel lines”. Nominal logistic regression does not assume parallel lines and can be used with categorical response variables which are not ordinal but have more than two categories, requiring the choice of a reference category.

*Online quizzes will be available for self-assessment.*
Course: STAA0010D - Multi-Level Linear Models

Blackboard Site: STAA0010

Dates: Wed 27/4, 4/5 (5:30pm – 8:30pm)

Room  ATC325

Assumed Knowledge: Multiple Linear Regression

Software used: SPSS and HLM7

Maximum 20 students

Course Description

Statistical techniques as listed below will be covered with an emphasis on the interpretation and reporting of these results.

When observations are clustered or auto-correlated conventional methods cannot be used. Such data is very common in practice and one of the advantages of these models is the way in which missing values can be handled. Mixed Linear Models are initially introduced to solve this problem. For more sophisticated problems HLM7 is a free student software package can be used. This software allows the fitting of longitudinal models and can handle response variables with a variety of distributions. Models are fitted separately for each subject and then combined to produce a population averaged model.

*Online quizzes will be available for self-assessment.*
Course: STAA0010E - Survival Analysis

Blackboard Site: STAA0010

Dates: Wed 11/5 (5:30pm – 8:30pm)

Room ATC325

Assumed Knowledge: Multiple Linear Regression

Software used: SPSS

Maximum 20 students

Course Description

Statistical techniques as listed below will be covered with an emphasis on the interpretation and reporting of these results.

Survival analysis considers the time until an event occurs and the chance of this even occurring at any time. If for any individual the event has not yet occurred the time must be censored, acknowledging that the actual time is longer the current time. Survival distributions for different treatments can be compared using a Kaplan-Meier Survival Probability Curve and log rank tests. Under certain conditions Cox Proportional Hazard models can be used to describe the risk of the event of interest occurring at any time. When these conditions are violated alternative models must be considered.

*Online quizzes will be available for self-assessment.*
Course: STAA0012 - Introduction to R

Blackboard Site: STAA0012

Dates: Thursdays 3/3 – 7/4 (18 hours)

Room: ATC325 5.30 – 8.30

Assumed Knowledge: Multiple Linear Regression

Software used: R

Course Description:

In this course you will learn how to install and configure R software. The course presents how to program in R, read data into R, access R packages, and organise and comment R code. In this course you will learn how to use R for effective analysis of the basic data types. Some of the most commonly used probability distributions will be introduced. Statistical data analysis will be conducted using working examples.

After successfully completing this unit, students will be able to:

1. Arrange and consolidate large datasets in an R software environment.
2. Develop the ability to perform advanced programing in an R software environment.
3. Relate the basics of fundamental probability distributions to different types of data.

Formulate practical and user friendly solutions to real life problems in the form of a statistical model in an R software environment.

Online quizzes will be available for self-assessment.