

## **Semester 1/Study Period 1 2022 Statistics Short Courses-Fully online (no on-campus classes)**

**Course: STAA0001 - Basic Statistics**

**Duration-12 weeks: from 28<sup>th</sup> Feb 2022 to 22<sup>nd</sup> May 2022**

**Canvas Site: STAA0001**

**Assumed Knowledge:** None

**Software used:** IBM SPSS Statistics Version 28

### **Course Description:**

This course develops students' understanding of a range of statistical methods along with their assumptions and limitations of their application. It enables students to develop the capacity to carry out independent statistical analysis of data using a standard statistical software package SPSS and also aims to develop students' abilities in effectively communicating the outcomes of statistical investigations.

Students who successfully complete this unit will be able to:

1. Effectively display information in datasets graphically.
2. Select an appropriate descriptive or inferential statistical technique based on the researcher's hypothesis, the level of measurement of the variables and testing of the appropriate assumptions to analyse the data.
3. Select appropriate IBM SPSS Statistics procedures, Java applets on the web and mathematical calculations, to obtain basic statistical test results, including confidence intervals and effect size statistics.
4. Explain the foundations of statistical inference, in particular the role of sampling distributions and the use of the normal distribution as a density curve.
5. Recognize when more advanced techniques are needed.
6. Write interpretive summary reports for both descriptive and inferential statistical analysis.

*Online quizzes will be available for self-assessment.*

**Course: STAA0002 - Simple Linear Regression and ANOVA**

**Duration-12 weeks: from 28<sup>th</sup> Feb 2022 to 22<sup>nd</sup> May 2022**

**Canvas Site: STAA0002**

**Assumed Knowledge: Basic Statistics (e.g. STAA0001**

**Software used: IBM SPSS Statistics Version 28**

**Course Description:**

This course of study aims to extend the ideas developed in Basic Statistics to include more advanced analyses, broaden the range of applications students are familiar with so that they will be able to carry out independent statistical investigations, and develop an awareness of the assumptions and limitations involved in the generalisation of results of such investigations.

Students who successfully complete this unit will be able to:

1. Choose the appropriate statistical analysis based on the researcher's hypothesis, the level of measurement of the variables and the testing of the appropriate assumptions.
2. Select appropriate statistical procedures, for example, using SPSS, Java applets on the web and mathematical calculations, to analyse data in a variety of contexts.
3. Relate the concepts of effect size, sample size, one or two tailed tests, level of significance and power of a statistical test.
4. Write interpretive summary reports for the inferential statistical techniques covered.
5. Judge when more advanced techniques are needed by comparing different statistical techniques for the variety of research questions.

*Online quizzes will be available for self-assessment.*

**Course: STAA0003A - Introduction to R**

**Duration-6 weeks: from 28<sup>th</sup> Feb 2022 to 10<sup>th</sup> April 2022**

**Canvas Site: STAA0003A**

**Assumed Knowledge: None**

**Software used: R and RStudio**

**Course Description:**

In this course participants will learn how to install and configure R software. They will also learn how to read data into R, access R packages, and organise and comment R code. Furthermore, they will learn how to use R for effective data analysis and visualisation. 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.
2. Develop the ability to perform basic programming in R.
3. Visualise data using R packages.
4. Relate the basics of fundamental probability distributions to different types of data.
5. Formulate practical and user-friendly solutions to real life problems in the form of a statistical model in an RStudio software environment.

*Online quizzes will be available for self-assessment.*

**Course: STA0003B-Using R for Statistical Analysis**

**Canvas Site: STAA0003B**

**Duration-6 weeks: from 11<sup>th</sup> April 2022 to 22<sup>nd</sup> May 2022**

**Assumed Knowledge:** Introduction to R (e.g. STAA0003A)

**Software used: R and RStudio**

**Course Description:**

This course of study aims to extend the ideas developed in Introduction to R. In this course you will learn key programming principles of R and how to develop and perform different types of statistical analyses and data visualisation. Participants expected to develop competence in programming in R - essential skill for a statistician or data scientist.

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

1. Write R programs to conduct hypothesis testing and compare means.
2. Perform simple linear regression in R.
3. Analyse categorical data using R.

*Online quizzes will be available for self-assessments*

**Course: STAA0006 – Introduction to SAS**

**Canvas Site: STAA0006**

**Duration: 6 weeks: from 28<sup>th</sup> Feb 2022 to 10<sup>th</sup> April 2022**

**Assumed Knowledge:** None

**Software used:** SAS

**Course Description:**

This unit is an introduction to one of the leading statistical software packages SAS - a powerful tool for organising and analysing data. Students will use this statistical software to develop their knowledge in data management, data presentation, and statistical analysis.

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

1. Demonstrate professionally relevant SAS programming and data management skills.
2. Write SAS programs to enter, load, and merge different types of data from multiple sources.
3. Visualise data graphically using SAS graphical procedures.
4. Summarise data and perform descriptive statistical analyses using SAS procedures.

*Online quizzes will be available for self-assessment.*

**Course: STAA0004A - Survey Design**

**Canvas Site: STAA0004**

**Duration: 6 weeks: from 28<sup>th</sup> Feb 2022 to 10<sup>th</sup> April 2022**

**Assumed Knowledge:** Basic Statistics (e.g. STAA0001)

**Software used:** None

**Course Description:**

Participants 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: STAA0004B - Research Design**

**Canvas Site: STAA0004**

**Duration: 6 weeks: from 11<sup>th</sup> April 2022 to 22<sup>nd</sup> May 2022**

**Assumed Knowledge:** Basic Statistics (e.g. STAA0001) **Software used:** Excel and IBM SPSS

Statistics Version 28

**Course Description:**

Participants will acquire skills and knowledge in observational and experimental studies, designing an experiment, incidence and prevalence statistics, different types of study designs including Cohort and Case-control studies.

**Topics will include**

- The basic concepts of experimental designs
- Common designs used in health statistics and elsewhere
- Incidence, prevalence and fertility statistics
- Mortality Statistics and Standardisation of rates
- Randomized trials and Cohort studies
- Case control studies

*Online quizzes will be available for self-assessment.*

**Course: STAA0005A - Multiple Linear and Logistic Regression**

**Canvas Site: STAA0005**

**Duration: 3 weeks: from 28<sup>th</sup> Feb 2022 to 13<sup>th</sup> March 2022 and 25<sup>th</sup> April 2022 to 1<sup>st</sup> May 2022**

**Assumed Knowledge:** Simple Linear Regression and ANOVA (e.g. STAA0002)

**Software used:** IBM SPSS Statistics Version 28 (R optional)

**Course Description:**

In Multiple Linear Regression we 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. Binary logistic regression will also be covered in this unit. **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.

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 and MANOVA**

**Canvas Site: STAA0005**

**Duration: 6 weeks: from 14<sup>th</sup> March 2022 to 24<sup>th</sup> April 2022**

**Assumed Knowledge:** Simple Linear Regression and ANOVA (e.g. STAA0002)

**Software used:** IBM SPSS Statistics Version 28 (R optional)

**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 analysis technique 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 factor analysis can then be used to 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).

Multi-variate analysis of variance (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.

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 - Advanced Statistical Modelling A: Generalised Linear Model**

**Canvas Site: STAA0010**

**Duration: 5 weeks: from 28<sup>th</sup> Feb 2022 to 3<sup>rd</sup> April 2022**

**Assumed Knowledge:** Multiple Linear Regression (e.g. STAA0005A)

**Software used:** IBM SPSS Statistics Version 28 and R

### **Course Description**

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

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.

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: STAA0010B - Advanced Statistical Modelling B:**

Mixed Models, Generalised Estimating Equations, Multi-level Models and Survival Analysis.

**Canvas Site: STAA0010**

**Duration: 5 weeks: from 11<sup>th</sup> April 2022 to 15<sup>th</sup> May 2022**

**Assumed Knowledge:** Multiple Linear Regression (e.g. STAA0005A)

**Software used:** IBM SPSS Statistics Version 28 and R

**Course Description:**

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

When observations are clustered or auto-correlated conventional methods cannot be used. Such data is very common in practise and one of the advantages of these models is the way in which missing values can be handled. Generalised Estimating Equations and 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.

Survival analysis follows. Kaplan Meier, Cox regression and Covariate Dependent Models.

*Online quizzes will be available for self-assessment.*

**STAA0013A The Basics of Scale Development**

**Canvas Site:** STAA0013

**Duration:** 6 weeks: from 28<sup>th</sup> Feb 2022 to 10<sup>th</sup> April 2022

**Assumed Knowledge:** Multiple Linear Regression (e.g. STAA0005A)

**Software used:** SPSS Version 28 and MPLUS

**Course Description:**

Following topics will be covered in this course:

- Introduction to types of scales and their development.
- Reliability.
- Validity.
- Exploratory factor analysis.
- Confirmatory factor analysis.

*Online quizzes will be available for self-assessment.*

## **STAA0013B Rasch Modelling**

**Duration:** 6 weeks: from 11<sup>th</sup> April 2022 to 22<sup>nd</sup> May 2022

**Canvas Site:** STAA0013

**Assumed Knowledge:** The Basics of Scale Development and Evaluation (STAA0013A)

**Software used:** IBM SPSS Statistics Version 28 and RUMM2030

### **Course Description:**

Following topics will be covered in this course:

- Rasch modelling. Lab: Rasch analysis using RUMM2030.
- Individual items and person analysis. Lab: Creating data file and analysis.
- Fit statistics, DIF and construct validity Lab: Analysis and interpretation of data.
- Dimensionality and scale targeting. Lab: Reading and critique articles.
- Issues in the use of scales in clinical and research settings; Lab: Develop and Evaluate a scale from a given data set.

*Online quizzes will be available for self-assessment.*