

## **Study Period 3 2024 Statistics Short Courses-Fully online (no on-campus classes)**

**Course: STAA0001 - Basic Statistics**

**Duration-12 weeks: from 26<sup>th</sup> Aug 2024 to 24<sup>th</sup> Nov 2024**

**Canvas Site: STAA0001**

**Assumed Knowledge:** None

**Software used:** IBM SPSS Statistics Version 29

### **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 26<sup>th</sup> Aug 2024 to 24<sup>th</sup> Nov 2024**

**Canvas Site: STAA0002**

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

**Software used: IBM SPSS Statistics Version 29**

**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 26<sup>th</sup> Aug 2024 to 6<sup>th</sup> Oct 2024**

**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 on 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 7<sup>th</sup> Oct 2024 to 24<sup>th</sup> Nov 2024**

**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 are expected to develop competence in programming in R - an 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: STAA0004A - Survey Design**

**Canvas Site: STAA0004**

**Duration: 6 weeks: from 26<sup>th</sup> Aug 2024 to 6<sup>th</sup> Oct 2024**

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

**Software used:** None

**Course Description:**

Participants will acquire skills and knowledge in the collection of a survey, observational, experimental, and secondary data, developing a questionnaire, and writing 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 7<sup>th</sup> Oct 2024 to 24<sup>th</sup> Nov 2024**

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

Statistics Version 29

**Course Description:**

Participants will acquire skills and knowledge in observational and experimental studies, designing an experiment, incidence, and prevalence statistics, and 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 Regression, General Linear Models and Multivariate Analysis of Variance**

**Canvas Site: STAA0005**

**Duration: 5 weeks: from 26<sup>th</sup> Aug 2024 to 29<sup>th</sup> Sep 2024**

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

**Software used:** IBM SPSS Statistics Version 29 (R optional)

**Course Description:**

In Multiple Regression, we will look at 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 and moderation using Process macro. Extensions of multiple linear regression including general linear modes (GLM) will also be covered in this course. 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 simple linear regression and ANOVA short course beforehand.

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

**Course: STAA0005B - Factor Analysis (EFA and CFA) and Introduction to Structural Equation**

**Modelling**

**Canvas Site: STAA0005**

**Duration: 6 weeks: from 30<sup>th</sup> Sep 2024 to 17<sup>th</sup> Nov 2024**

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

**Software used:** IBM SPSS Statistics Version 29 (R optional) and AMOS

**Course Description:**

Factor Analysis covers both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). In the 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 scores.

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, with 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. Introductory structural equation modelling (SEM) techniques will also be covered during the last two weeks of this course. SEM helps us to investigate patterns of effect within a system of variables which can examine the impact of a set of predictor variables on multiple dependent variables.

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



**Course: STAA0010A – Models for Categorical Variables**

**Canvas Site: STAA0005**

**Duration: 6 weeks: from 26<sup>th</sup> Aug 2024 to 6<sup>th</sup> Oct 2024**

**Assumed Knowledge:** Multiple Linear Regression (e.g., STAA0005)

**Software used:** IBM SPSS Statistics Version 29 (R optional)

**Course Description:**

This unit of study covers a wide range of linear models. In particular it considers

- a) The use of categorical independent variables in general linear models
- b) Binary logistic regression models with binary dependent variables
- c) Ordinal logistic regression models with ordinal dependent variables
- d) Nominal logistic regression models with nominal dependent variables with three or more categories
- e) Various methods for the imputation and management of missing data are also covered.

These methods are commonly used for the analysis of survey, clinical trial and other experimental data. In particular, binary logistic regression is commonly used for the analysis of attrition in clinical trials, non-completion of surveys and for identifying any demographic or baseline differences between treatment groups.

Weekly Online quizzes will be available for self-assessment

**Course: STAA0010B** – Generalised Linear Models for Cross-sectional, Longitudinal and Survival Data

**Canvas Site: STAA0005**

**Duration: 6 weeks: from 14<sup>th</sup> Oct 2024 to 24<sup>th</sup> Nov 2024**

**Assumed Knowledge:** Multiple Linear Regression (e.g., STAA0005)

**Software used:** IBM SPSS Statistics Version 29 (R optional)

**Course Description:**

This unit of study covers a wide range of linear models. In particular it considers models for

- a) Cross-sectional data (Generalised Linear Models)
- b) Longitudinal data (Generalised Estimating Equations, Generalised Linear Mixed Models)
- c) Survival data (Kaplan-Meier Analysis and Cox Regression)

It also considers a variety of distributions for the dependent variables. In addition to normal distributions it considers

- a) Poisson for count data, but allowing for over-dispersion when the variance exceeds the mean
- b) Negative Binomial for more variable count data
- c) Binomial for the number of successes in n trial, again allowing for over-dispersion when necessary
- d) Survival/Reliability studies.

These methods are commonly used for the analysis of observational or experimental studies, allowing the comparison of treatment and control (or pseudo-control) groups at baseline and over time. Survival models are used to identify risk factors for mortality, disease, relapse and mechanical failure.

Weekly Online quizzes will be available for self-assessment