Study Period 3- 2023 Statistics Short Courses-Fully online (no on-campus class)

These short courses will be conducted through Hawthorn Online (HOL) teaching calendar and will begin on Monday 28th August 2023.

Course: STAA0001A - Basic Statistics

Duration-12 weeks: from 28th August 2023 to 26th November 2023

Canvas Site: STAA0001A

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 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 28ᵗʰ August 2023 to 26ᵗʰ November 2023

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 28th August 2023 to 8th October 2023

Canvas Site: STAA0003A

Assumed Knowledge: None

Software used: 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.
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: STAA0003B - Using R for Statistical Analysis

Canvas Site: STAA0003B

Duration: 5 weeks: from 9th October 2023 to 19th November 2023

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

Software used: 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.
3. Analyse categorical data.

Online quizzes will be available for self-assessment.

Course: STAA0004A - Survey Design

Canvas Site: STAA0004A

Duration-6 weeks: from 28th August 2023 to 8th October 2023
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: STAA0004B

Duration: 6 weeks: from 9th October 2023 to 26th November 2023

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, 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 and Multivariate Analysis of Variance (MANOVA)

Canvas Site: STAA0005A

Duration-5 weeks: from 28ᵗʰ August 2022 to 1ˢᵗ October 2023
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 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. 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 and Introduction to Structural Equation Modelling (SEM)

Canvas Site: STAA0005B

Duration: 6 weeks: from 2ⁿᵈ October 2023 to 26ᵗʰ November 2023

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

Software used: IBM SPSS Statistics Version 29 and AMOS 29

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. Introductory structural equation modelling (SEM) techniques will also be covered during 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.

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

Canvas Site: STAA00010A

Duration-11 weeks: from 28th August 2023 to 26th November 2023

Software used: SPSS, R and R studio

Pre-requisites: Multiple Linear Regression (e.g. STAA0005A) and MANOVA (e.g. STAA0005B)

This short course considers data which does not comply with the usual regression assumptions, while also managing missing data. In particular, it considers dependent which do not have normal distributions (e.g. categorical variables and count variables). This includes binary, ordinal, multinominal and loglinear regression models and Generalized linear models (GLM). In addition the short course considers time-to-event analyses, sometimes referred to as Survival Analysis. Finally, repeated measures models are considered that allow for complicated data structures while also allowing for missing data. These methods include Generalized estimating equations (GEE) for large numbers of observations, Mixed Models and Multi-level modelling for more sophisticated unit-specific modelling.

Online quizzes will be available for self-assessment.

Course: STAA0011A –Advanced Data Mining

Canvas Site: STAA00011A

Duration-11 weeks: from 28th August 2023 to 19th November 2023
Software used: R and R studio and the Rattle interface

Pre-requisites: Multiple Linear Regression and MANOVA (e.g. STAA0005A) and Factor Analysis

Course Description:

This short course introduces the principles and techniques used by data miners using R and the Rattle interface. Data mining is used to add value to large collections of data, delivering discoveries that continue to revolutionarise lives in our data-rich but knowledge-hungry world.

In particular, this short course introduces classification methods such as Classification Trees, Random Forests, Boosting, support vector machines and Neural Networks. In addition, it considers modern variable selection methods for linear models, and non-linearity using splines. Finally, cluster analysis for creating homogeneous and market basket analysis for association analysis are introduced, with a very brief introduction to text analysis.

A very practical approach is adopted throughout the short course with an emphasis on implementation and the strategic use of these methods in any industry, often in the context of "big data".

Online quizzes will be available for self-assessment.