

# DATA ANALYSIS WITH R

*This module provides the learner with advanced training in statistical methods with R, relevant to data analysis in the design and planning of experiments as part of the research process. It will address statistical methodologies and applications to research.*

## LEARNING OUTCOMES:

On successful completion of this module the learner will be able to:

- Use statistics to reduce complex data situations to manageable formats in order to describe, explain or model them.
- Derive descriptive statistics for various data types using R.
- Perform and critique statistical tests on two sample data using R.
- Set up and critically analyse data sets in both a parametric and non-parametric way for two and more samples using R.
- Use multiple regression with R and other advanced statistical techniques to allow prediction of a score on one variable on the basis of the scores on several other variables.
- Communicate effectively research findings in a clear concise manner using correct terminology based on output from R.

## INDICATIVE CONTENT:

### Overview

Statistics fills the crucial gap between information and knowledge. Society cannot be run effectively on the basis of hunches or trial and error. This topic highlights which statistics to use, why to use those statistics, and when to use them.

### Introduction to data analysis

Through using appropriate descriptive statistics, it is possible to make sense of data collected and tell a research story coherently and with justification. This entails deriving the correct measures of centrality/variation, if applicable. Along with interpreting bar charts, pie charts, histograms, stem-and-leaf plots, boxplots.

### Statistical inference (two samples)

Understand the idea behind hypothesis testing through worked examples of test of normality/differences/relationships with various types of data - i.e., Independent and related t-tests; Mann Whitney and Wilcoxon test; Pearson and Spearman Rank correlation.

### Multi-variable analysis

Set-up and analyse various data sets in both a parametric and non-parametric way. In the case of non-parametric data, suitable data transformations will be investigated prior to the use of parametric tests. One-way Analysis of Variance (ANOVA) with suitable posthoc testing. Between and within subjects factorial experiments. Investigation of the effect size of a result and the power to a test.

## MODULE RESOURCES

### Recommended Book Resources

- Tony Fischer 2015, Data Analysis with R, Packt Publishing [ISBN: 978-178528814]
- Hadley Wickham, Garrett Golemund 2016, R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, 1 Ed., O'Reilly Media [ISBN: 1491910399]
- Peter Dalgaard 2008, Introductory Statistics with R, Springer New York [ISBN: 9780387790534]

### Supplementary Book Resources

- Michael J. Crawley 2012, The R Book, Wiley-Blackwell [ISBN: 978-047097392]
- Andrew P. Beckerman, Dylan Childs and Owen Petchey 2017, Getting Started With R, Oxford University Press [ISBN: 978-019878784]
- Tadhg L. O'Shea 2013, Essential Statistics for Researchers [ISBN: 978-0-9575059-0-2]

### Other Resources

- Website: Data Camp:  
<https://www.datacamp.com/>

### Multiple regression

Scatterplots and partial regression plots. Test for homoscedasticity. Detect for multicollinearity and outliers. Check that the residuals (errors) are approximately normally distributed. Interpret regression equations and use them to make predictions.

### R as a statistical programming language

R will be used to turn raw data into insight, knowledge, and understanding. Packages such as: ggplot2 will be used to declaratively create graphics; dplyr for data manipulation; tidyr for a set of functions that help you get to tidy data; readr will be used to provide a fast and friendly way to read rectangular data; purrr for providing a complete and consistent set of tools for working with functions and vectors.

## COURSE WORK

Assessment Type	Assessment Description	Outcome addressed	% of total	Assessment Date
Practical /Skills Evaluation	Use the appropriate descriptive statistics tools in R to suggest whether a difference exists between the two measurements. Summarise all results in a concise report	1,2,6	20.0	Week 5
Practical /Skills Evaluation	Using functions in R, test the significance of hypotheses that differences exist between two measurements. Summarise results in a concise report.	3,6	30.0	Week 7
Practical /Skills Evaluation	Using R, test the significance of a hypothesis that a difference exists between more than two measurements. Apply posthoc tests to determine the presence of statistical differences. Summarise results in a concise report.	4,6	30.0	Week 10
Practical /Skills Evaluation	Using the appropriate function in R, test the relationship between a response variable and multiple explanatory variables. Summarise results in a concise report	5,6	20.0	Sem End