www.academicdataanalysts.org

Essential Level

Online Academic Data Analysis Bootcamp Using Open-Access Program R

About the course

R is broad and powerful, with many analytic and graphic functions available (more than 50,000). With guidelines and instructions, you can navigate the tremendous resources available in R thereby accomplishing your work with style, elegance and efficiency.

Many people and researchers despise statistics, mainly due to non-mathematical background. This makes understanding complex statistical equations very difficult. The advent of computer programs such as R and the likes provide a unique opportunity to teach statistics at a conceptual level without getting too bogged down in equations. However, the downside of the computer is that it makes it really easy to make complete fools of ourselves if we do not really understand what we are doing. Running analysis using a computer without any statistical knowledge can be a totally misleading. Hence this course could be called *Unearthing the Statistician in You Using R*.

The course is split into 3 levels, i.e. essential, intermediate, and advanced topics, each taking the same time as the previous combined course. This will allow for slow organic tutoring, beginning with the basic structures of the R codes and key statistical concepts leading to a discovery journey that will allow you to quickly access the power of this great open source endeavor. This course attempts to strike a good balance between theory and practice by using the computer as a tool for learning statistical concepts with the hope that you will gain a better understanding of both theory and practice. A **Certificate of Completion** will be issued upon the successful completion of the course.

Course Content

Levels	Sessions	Modules
Essentials	**Session 1**	**1. The R environment** 1.1 Why R, and R Paradigm 1.2 References, Tutorials and links 1.3 R Overview 1.4 R Interface, Rstudio, R Script 1.5 Commands, objects and functions 1.6 Points to note 1.7 R Working directory and Workspace 1.8 Help 1.9 R-chitecture and R Packages 2.0 Data types

Enhancing access to world-class data analysis skills and open access tools to less reached groups

	Session 2	**2. Getting data into and out of R** 2.1 Creating variables and data frames 2.2 Calculating new variables from existing ones 2.3 Working with dates 2.4 Missing values 2.5 Entering data, creation of variables and coding variables using R Commander 2.6 Using other software to enter and edit data 2.7 Importing data 2.8 Exporting data 2.9 Viewing and Examining data
	Session 3	**3. Data Management in R** 3.1 Sorting Data 3.2 Merging Data 3.3 Aggregating Data 3.4 Reshaping Data 3.5 Subsetting Data 3.6 Data Type Conversion
	Session 4	**4. Plotting in R** 4.1 Creating a Graph 4.2 Histograms and Density Plots 4.3 Dot Plots 4.4 Bar Plots 4.5 Line Charts 4.6 Pie Charts 4.7 Boxplots 4.8 Scatterplots 4.9 Advanced Graphics: graphical parameters, axes and text, combining plots
	Session 5	**Part 1** **3 Exploring assumptions** 3.1. What will this chapter tell me? 3.2. What are assumptions? 3.3. Assumptions of parametric data 3.4. Packages used in this chapter 3.5. The assumption of normality 3.5.1. Oh no, it's that pesky frequency distribution again: checking normality visually 3.5.2. Quantifying normality with numbers 3.5.3. Exploring groups of data 3.6. Testing whether a distribution is normal

		3.6.1. Doing the Shapiro–Wilk test in R 3.6.2. Reporting the Shapiro–Wilk test 3.7. Testing for homogeneity of variance 3.7.1. Levene's test 3.7.2. Reporting Levene's test 3.7.3. Hartley's *F*max: the variance ratio **Part 2** 3.8. Correcting problems in the data 3.8.1. Dealing with outliers 3.8.2. Dealing with non-normality and unequal variances 3.8.3. Transforming the data using R 3.8.4. When it all goes horribly wrong F
	Session 6	**Part 1** **4 Correlation** 4.1. What will this session tell me? 4.2. Looking at relationships 4.3. How do we measure relationships? 4.3.1. Standardization and the correlation coefficient 4.3.3. The significance of the correlation coefficient 4.3.4. Confidence intervals for *r* 4.3.5. A word of warning about interpretation: causality 4.4. Data entry for correlation analysis 4.5. Bivariate correlation 4.5.1. Packages for correlation analysis in R 4.5.2. General procedure for correlations using R 4.5.3. Pearson's correlation coefficient 4.5.4. Spearman's correlation coefficient 4.5.5. Kendall's tau (non-parametric) 4.5.6. Bootstrapping correlations 4.5.7. Biserial and point-biserial correlations **Part 2** 4.6. Partial correlation 4.6.1. The theory behind part and partial correlation 4.6.2. Partial correlation using R

		4.6.3 Semi-partial (or part) correlations 4.7. Comparing correlations 4.7.1. Comparing independent *r*s 4.7.2. Comparing dependent *r*s 4.8. Calculating the effect size 4.9. How to report correlation coefficients Interesting real research

The Tutors

About the Lead Tutor

Dr. Patrick Njage is a Statistician working as a Genomic Epidemiologist at the National Food Institute, Technical University of Denmark. His passion and ability to teach both common and advanced statistical topics with great simplicity in open access language R stems from his background in two worlds. He has background training in both natural sciences (BSc, MSc, PhD (Microbiology and Biotechnology) as well as Statistics (Master of Statistics, University of Hasselt, Belgium). He is experienced in the analysis of complex, unstructured quantitative data. He has taught statistical analysis in R environment to students in funded fellowships from 5 African countries and workshops as well as lectures to participants from different countries in Europe. He has also supervised and co-supervised numerous under and post graduate student projects.

Joshua Ombaka(Ph.D), African Centre for Technology Studies

Faith Muniale (Ph.D.) Tropical Biology Association

Parita Shah (Ph.D.) Department Of Geography and Environmental Studies, University Of Nairobi, Kenya

Patrick Njage, Ph.D Technical University of Denmark

Enhancing access to world-class data analysis skills and open access tools to less reached groups

www.academicdataanalysts.org

Enhancing access to world-class data analysis skills and open access tools to less reached groups