

## Data Analysis Bootcamp Using Open-Access Program R (Essential Level)

### About the course:

R is broad and powerful, with many analytic and graphic functions available (more than 20,000). 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 like provides 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 easy to err critically if we do not understand what we are doing. Running an analysis using a computer without any statistical knowledge can be really misleading. Hence this course could be called *Unearthing the Statistician in You Using R*.

The course is the first of 3 levels, i.e. Essential, Intermediate, and Advanced levels. This will start you off using 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	<b>1. The R environment</b> <ul style="list-style-type: none"> <li>1.1 Why R, and R Paradigm</li> <li>1.2 References, Tutorials and links</li> <li>1.3 R Overview</li> <li>1.4 R Interface, RStudio, R Script</li> <li>1.5 Commands, objects and functions</li> <li>1.6 Points to note</li> <li>1.7 R Working directory and Workspace</li> <li>1.8 Help</li> <li>1.9 R-architecture and R Packages</li> <li>2.0 Data types</li> </ul>
	Session 2	<b>2. Getting data into and out of R</b> <ul style="list-style-type: none"> <li>2.1 Creating variables and data frames</li> <li>2.2 Calculating new variables from existing ones</li> <li>2.3 Working with dates</li> <li>2.4 Missing values</li> </ul>

Levels	Sessions	Modules
		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
	<b>Session 3</b>	<b>3. Data Management in R</b>  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
	<b>Session 4</b>	<b>4. Plotting in R</b>  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
	<b>Session 5</b>	<b>Part 1</b>  <b>3 Exploring assumptions</b> 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

Levels	Sessions	Modules
		<p>3.7.1. Levene’s test 3.7.2. Reporting Levene’s test 3.7.3. Hartley’s <math>F_{max}</math>: the variance ratio</p> <p><b>Part 2</b></p> <p>3.8. Correcting problems with 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</p>
	<p><b>Session 6</b></p>	<p><b>Part 1</b></p> <p><b>4 Correlation</b> 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 <math>r</math>     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</p> <p><b>Part 2</b></p> <p>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</p>

Levels	Sessions	Modules
		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

**What to expect:**

R is powerful, versatile, and easy to use statistical software. It is the first choice for many data analysts working in both companies and academia. The course will help you master the basics of R in a short time, as a first step to becoming a skilled R data analyst. The course is suitable for beginners with absolutely no knowledge in R as well as intermediate users and refresher participants.

By the end of the course, you will have the most important R programming skills to handle data analysis from data cleaning and exploration to the inferential analysis of various levels of complexity. Also, you will be able to identify appropriate analysis for your type of research problem. Additional materials to help you understand the statistical software will be distributed during the course.

**Requirements**

No special prerequisite except basic high school-level mathematics. However, you require passion to learn and need basic knowledge on how to use a computer/ laptop. Also, you can prepare your personal data for research application of the concepts learned. You are welcome to bring your own data analysis questions to the resource persons.

**Course Format**

The course will use interactive teaching techniques in breakout groups, and each participant will be encouraged to participate based on his or her own field of research, work or study. We hope that each participant will bring with them a case study on research they have been involved. Example datasets will also be provided for each concept. In addition to learning and skills development, the course will provide a valuable opportunity for those involved in quantitative data analysis to meet and later join an R programming support community consisting of the course alumni.

The sessions will be structured as follows:

- 30- 60 minutes introductory lecture (Tuesdays prior to the practical session): introduction to the statistical concepts and practical session of the week
- 2 hours: guided hands-on practical session and discussions with exercises given in the class.

### **Registration**

The course fee for this course ranges between USD450 (KES 58,315.50) - USD500 (KES 64795.00). However, to achieve the vision of enhancing access to world-class data analysis skills by the “less reached groups in Africa”, the course does not include a tuition fee and only a registration fee of (KES 5,000) required to support maintenance of the online study portal, award of certificates and other expenses including honoraria for the volunteer instructors and moderators.

### **Registration fee payment**

Registration will be via KEMRI Graduate School website available here: <https://kgs.ac.ke/short-course-programmes/> by clicking on the “**Data Analysis Bootcamp Using Open-Access Program R**”

Registration deadline is on **19<sup>th</sup> April 2026**. Payment deadline is **26<sup>th</sup> April 2026**. The course will run for 6 weeks from **5<sup>th</sup> May 2026; Tuesdays (7:00–8:30 PM) and Thursdays (7:30–9:30 PM)**.

### **Pre-course Zoom orientation (Q&A session)**

Prior to the registration deadline, we have organized a pre-course Zoom orientation and Q&A session on **Thursday 16<sup>th</sup> April, 2026 at 19:30 EAT**. You are welcome to attend this session via the Zoom link below. A Certificate of Completion will be provided upon successful completion of the course.

**Zoom meeting link for pre-course Zoom orientation and Q&A session:** <https://shorturl.at/91I1Q>

Meeting ID: 623 3920 7277

Passcode: 177224