Statistics and Environmental Analytics 2020 [Online Course]

Module Guide

Dr. Rohan Shetti (PhD)

Landscape Ecology and Ecosystem Dynamics

Greifswald University Germany.

Course background Information:

This course is designed for students from all disciplines within Environmental and Earth Sciences. The main aim of this module is to help researchers, students and professionals to apply statistical concepts to their data using R programming. The module assumes that the people attending have at least a basic understanding of statistics and have previously worked with some or the other form of environmental data. In this module, we aim to brush up basic concepts from statistics and apply them to real environmental data. We aim at providing the attendees with the critical knowledge of how to determine which statistical test are relevant to their data-set and how to apply those using R and R Studio programming interfaces. Additionally, we aim to cover how different outcomes can be presented using various plots and diagrammatic representations. Finally, we would also like to invite students to bring samples from their data-sets in case they need any help or directions in their analysis. For more information about the module kindly visit the website www.theenvironmentalanalytics.com

This course **IS** for you if you are a

1. Student/ researcher/ professional who has a basic understanding of statistics but wants to learn applied R / R Studio programming for dissertation/ research analysis.
2. Have a background understanding of statistics but are not sure how to determine which tests I should be using for my data.
3. Have a basic understanding of R programing language but needs help with translating statistical tests on real-time data.
4. To upscale your concepts/algorithms to make them sharable with other researchers.
5. Wish to learn how to publish your R scripts on internet portals to share your algorithms with colleagues, co-workers and other scientists in your discipline.
6. Wish to understand and interpret the graphical representation of statistical inferences in articles, journals and reports.

This course is **NOT** for you if:

1. You are already an advanced R programmer with thorough knowledge of statistics.
2. Have a sound stats background and don't need to use programming for analysis.
3. If you don't require data-science in your discipline of work.

Day 1: Introduction to Analytics and Ecological context

Pre-lunch - Inculcating a modellers mind-set (2 hours)

1. Interactive session: Keyword deduction of environmental processes.
2. Interactive session: i. Dynamic interrelations in the environment.
   ii. From functional observations to statistics and programming.
3. Analytics in an Ecological context (presentation with interaction)
4. Introduction to data processing platforms: R / R-Studio

LUNCH Break for 1hr 30 mins

Post-lunch – Installation and introduction to R (2 hours)

1. Installation of R and R-Studio on student machines
2. Introduction to R graphical user interface.
3. Basic operations in R.
4. Data cleaning and import.
5. Setting homework

Day 2: Basic operations in R, data-import and structuring and analysis

Part 1: 2 hours

1. Data story: Dendro-data from the Arctic: Explaining the "limiting factor" principle and setting the premise of an analytical story.
2. Importing data (provided by me) and extracting data from online sources; Climate Explorer using GPS co-ordinates and importing data in R.
3. Data viewing: overview, and how to understand data (principle of biotic vs abiotic in a matrix framework) (demo and presentation)

LUNCH BREAK 1hr 30 mins

Part 2: 2 hours

4. Data cleaning; looking for NAs, 0, missing values (explaining the inter-relations).
5. Data plotting; simple line plots.

Day 3: Statistics and Data representation in operations in R

Case Study:

Pre-lunch

1. Introduction to loops.
2. Activity for students to write loops and plot
3. GGplot2

LUNCH BREAK 1hr 30 mins

Post-lunch

1. Introduction to functions.
2. Example of functions and activity for students to write simple functions.
3. Ready functions for large data sets (*"mapply" "apply"* and *"tapply"*)

Closing session:

Questions about students data-sets or general questions about the course and feedback.