



SHORT COURSE TRAINING ON “*STATISTICAL METHODS FOR MANAGEMENT, ANALYSIS, AND VISUALIZATION OF BIG DATA WITH R AND MS. EXCEL*”

1. Introduction

As our daily activities expand, a significant volume of data is generated. Merely seeing this data in isolation is insufficient for making well-informed decisions without a comprehensive understanding and analysis by a data analyst. Due to the recent advancements in advanced techniques and software for data analysis, it has become more convenient to analyze and visualize the patterns within large data sets.

This short course is designed to equip learners with various statistical methods for analyzing and visualization techniques to make sense of the emergence and exponential growth of big data. This will in turn lead to better decision-making, insights and advantages of having such big data. Therefore, our main aim is to help learners understand fundamental concepts of big data analytics, and explore and exhaust every information hidden in the big data. The course will be delivered mainly through lectures, discussions and demonstrations. During the course, R and MS Excel statistical software will be used to analyze and visualize big data.

2. Course Objective

This short course intends to equip participants with the following knowledge and skills;

- (i) Differentiate traditional and Big Data datasets.
- (ii) Storage, management and analysis of both structured and unstructured data sets.
- (iii) Apply principles of statistical analytics to solve problems and inform decision-making.

3. Course Outcomes

At the end of this course, participants will be able to;

- (i) Differentiate traditional and big datasets.
- (ii) Design different types of data analysis tasks.
- (iii) Conceptualize and design different types of data analysis tasks.
- (iv) Interpret the outputs of data analysis projects.

4. Course Outline

The following topics will be covered during the training.

Topic 1: Introduction to Big Data Analysis

Big data concept, traditional and big datasets, Data collection, cleaning, transformation & integration, streaming data analysis, structured data analysis, social network analysis and Text analysis.

Topic 2: Introduction to R and Excel Basics

Part 1: Excel Basics

Data processing and cleaning, sorting, filter and removing duplicates, useful data manipulation formulas and functions, use pivot tables and slicer to describe data, charts/graphs and tables. Dashboard design.

Part 2: R Basics

Installing and working with R, data structures, basic arithmetic operations, measures of central tendency and dispersion, Data frames, slicing and indexing, basic plots and graphics, importing/exporting data to/from R, basic loops and control statements, Simple functions.

Topic 3: Basics of Data Analysis

Understanding of data dimensions, volume, variables and structure, five number summary. Exploratory data analysis (EDA) through formulating research question/objective, try easy solution(s) and solution(s) challenge. Exploratory graphs/tables, etc.

Topic 4: Statistical Methods for Analyzing Big Data.

Data manipulation techniques, Divide and conquer methods, sampling methods, linear regression and prediction.

Topic 5: Data Visualization and Reporting the Results

Plotting systems for qualitative and quantitative data. Customize plots through appropriate labels, headings and legends. Interpretation of plots/graphs as well as tables created. Saving the plots and tables for further analysis. How best to report the results to be meaningful and easily understood to the general audience. The dos and don'ts in discussion of results/tables/graphics.

5. Participation Fee and Mode of Payment

A participant is required to pay Tshs 1,000,000/= (One Million) which covers meals and training materials. Payment should be made through account number 20110016259, account name: TASTA, NMB Bank.

6. Place of Training

The short course will be conducted at Dodoma City from 10th to 18th June 2024.

7. Confirmation of Attendance

Participants who wish to attend the training should confirm their attendance by 7th June 2024 by sending payment slip through info@tastastat.or.tz.