Introduction to Statistics and Data Science

<u>Time</u>: 12:00 pm – 2:00 pm GMT, every SATURDAY
<u>Course Span</u>: 30th August 2025 – 13th December 2025
<u>Course Link:</u> *to be updated*
<u>Course Survey Link</u>: [t.ly/Z6cut](t.ly/Z6cut)

Instructor: Charles Opondo   <u>Office Hours</u>: By appointment
E-mail: charlesopondo1376@gmail.com

Instructor: Ruchir Raman
Email: ruchinbox@gmail.com

Instructor: Habibah Tijani
Email: tijanijumoke410@gmail.com

Course Overview
This course provides a foundational introduction to statistics, focusing on essential concepts and practical applications using Microsoft Excel and R. It is designed for undergraduate students and above with limited or no background in statistics.

Course Objectives
By the end of this course, students will be able to:
- Understand fundamental statistical concepts such as descriptive statistics, probability, and hypothesis testing
- Apply basic statistical methods to analyze data.
- Use Excel and R for statistical computations and visualization.
- Interpret statistical results and communicate insights effectively.

Course Structure
This course follows a modular, hands-on approach to learning statistics and data science principles. The structure alternates between lecture sessions that introduce statistical concepts and lab sessions that reinforce learning through real-world applications in Microsoft Excel and R.
- **Lecture Sessions**
  Lectures will introduce statistical concepts, contextualized within real research examples. They will also include demonstrations and Q&A to prepare students for hands-on application.
- **Lab Sessions**
  Labs will apply concepts from the previous week using datasets. Labs 1–3 will use

Excel to build foundational skills, after which students will **transition to R** for more advanced statistical methods starting from Week 8.

- **Assignments**
  Bi-weekly tasks will reinforce lecture and lab content, covering both conceptual understanding and practical applications.
- **Final Project**
  Students will conduct a basic data analysis project from start to finish. They will apply descriptive statistics, hypothesis testing, and simple modeling using R, and present findings in both written and visual format.
- **Office Hours & Support**
  Office hours available by appointment. Support is also available via email for technical or conceptual help.
- **Attendance**
  Students are required to attend 80% of the classes and submit all bi-weekly assignments and the final project to be eligible for the STEM for development certificate.

Course Evaluation

Students are required to attend 80% of the classes and submit all bi-weekly assignments and the final project to be eligible for the STEM for Development certificate.

Weekly Course Schedule and Meeting Information
Time: 12:00 pm – 2:00 pm GMT
Course Link: *to be updated*
Course Survey Link: t.ly/Z6cut

WEEK	DATE	TOPIC	INSTRUCTOR	DESCRIPTION
1	30th August 2025	Introduction to Statistics & Data Types	Ruchir + Charles	a. Understanding statistical concepts and applications. b. Overview of categorical, ordinal, and numerical data. c. Introduction to Excel and R for statistical analysis.
2	6th September 2025	Lab – Working with Data in Excel	Habibah	a. Importing and organizing datasets in Excel. b. Identifying different data types.

				c. Basics of Exploratory Data Analysis (EDA).
3	13th September 2025	Descriptive Statistics	Charles	a. Measures of central tendency (mean, median, mode). b. Measures of dispersion (variance, standard deviation, range). c. Data visualization (histograms, charts, tables and box plots).
4	20th September 2025	Lab – Descriptive Statistics in Excel	Habibah	a. Computing summary statistics. b. Creating visualizations (histograms, box plots). c. Interpretation of results.
5	27th September 2025	Probability Basics	Ruchir	a. Introduction to probability. b. Common distributions (normal, binomial). c. The Law of Large Numbers.
6	4th October 2025	Lab – Probability Applications in Excel	Habibah	a. Simulating probability distributions. b. Calculating probabilities using Excel. (Final Excel-Based Lab)
7	11th October 2025	Sample Surveys	Charles	a. Understanding probability and non-probability sampling techniques. b. Advantages and disadvantages of probability and non-probability sampling techniques

				c. Sources of errors in surveys d. Comparing methods of data collection in surveys.
8	18th October 2025	Lab – Intro to R & Data Import	Habibah	a. Installing R and RStudio. b. Importing and viewing datasets. c. Navigating basic R syntax. *(First R-based lab)*
9	25th October 2025	Introduction to inferential statistics	Ruchir	a. Hypothesis testing • Null and alternate hypothesis • Statistical significance and p-value • Power concepts b. Confidence intervals c. Central limit theorem
10	1st November 2025	Fundamental statistical tests	Charles (parametric) Ruchir (non-parametric) Both (ANOVAs)	a. Tests for parametric data • t-test (independent and paired) • z-test b. Tests for non-parametric data • Chi-square test • Mann-Whitney U test • Wilcoxon Sign and Rank tests • McNemar test c. Analysis of Variance (ANOVAs) • Multiple corrections problem

				● Post-hoc tests d. Choosing the best statistical test
11	8th November 2025	Lab – fundamental statistical tests and hypothesis testing	Habibah	a. Work with real/simulated data for concepts. b. Interpretation of hypothesis related concepts (significance, rejection/failure of rejection of null hypothesis, confidence intervals)
12	15th November 2025	Conceptual correlation and regression	Charles	a. Understanding correlation b. Scatterplot c. Identifying relationships d. Introduction to simple linear (*model as a line*) regression and logistic regression. e. Using regression analysis in research.
13	22nd November 2025	Lab – Correlation and Regression in R	Habibah	a. Use correlation functions b. Visualize patterns. c. Build basic linear and logistic regression models d. Model output interpretation
14	29th November 2025	Final Project Overview (shortened session: 1 hour)	All 3	a. Project guidelines, group/team selection.

15	6th December 2025	Final Project Work Session (Optional)	All 3	a. Peer support and instructor troubleshooting.
16	13th December 2025	Project Presentations & Course Wrap-up	All 3	a. Presentation of final projects. b. Course review and feedback. c. Certification eligibility review.

Software Requirement
- Microsoft Excel (2016 or later)
- R (Core Programming Language)
- R-Studio (User Interface)

Step 1: Download and install **R**
https://cran.r-project.org

Step 2: Download and install **RStudio Desktop (free)**
https://posit.co/download/rstudio-desktop

Required textbook
Madsen, B. (2016), *Statistics for Non-Statisticians*. 2nd ed. Springer.

Learning Resources

- *R for Data Science* (2e) – Free online: r4ds.hadley.nz
- *Handbook of Biological Statistics* – biostathandbook.com
- *OpenIntro Statistics* – Free textbook: openintro.org
- *Statistics Using Excel* – Khan Academy: https://www.khanacademy.org/math/statistics-probability
- *Introduction to R* – DataCamp: https://www.datacamp.com/courses/free-introduction-to-r
- *StatQuest Videos* – https://www.youtube.com/user/joshstarmer
- *The Art of Statistics* – Spiegelhalter *(Suggested reading for insight & fun)*