



Statistics for Data Science

(Website: www.aiquest.org, Cell: +8801704265972)

About Course:

Course Duration: 40 to 42 Hours

Total No. Of Classes: 18-20

Class Per Week: 02 at 8:30 - 10:30 P.M.

Class Duration: 2 Hours (Approx)

Tools & Languages: R, SPSS, PowerBi

Course Fees: 5000 Taka / 50\$

Course Instructor:

Mr. Zarin Hasan

Business Intelligence Analyst at Apple Gadgets Ltd.

Bachelor in Statistics, Islamic University, Kushtia

Module 01: Introduction to Statistics for Data Science (2 Hours)

- ❖ The Importance of Statistics:
- ❖ Statistics in Data-Driven World:
- ❖ Future of Statistics:
- ❖ Interdisciplinary Applications:
- ❖ Statistical Exploration and Visualization:
- ❖ Data-Driven Decision Making:
- ❖ Ethics and Responsibility:
- ❖ Career Opportunities:
- ❖ Continuous Learning and Adaptation:
- ❖ Practical Application vs. Theory:
- ❖ Your Learning Journey:

Module 02: R Programming for Data Science (6 Hours)

Session 1: Introduction to R Basics (1 hour)

1. Introduction to R:
 - History and purpose of R.
 - Installing R and RStudio.
2. RStudio Interface:
 - Overview of the RStudio layout.
 - Exploring the console, script editor, environment, and plots panel.
3. Basic Operations:
 - Arithmetic operations.
 - Assigning variables.
 - Data types: numeric, character, logical.
4. Vectors:
 - Creating and manipulating vectors.
 - Vectorized operations.

Session 2: Data Structures and Control Flow (2 hours)

1. Matrices and Arrays:
 - Creating matrices and arrays.
 - Matrix arithmetic.
2. Lists:
 - Creating and working with lists.
 - Accessing list elements.
3. Data Frames:
 - Creating data frames.
 - Subsetting and manipulating data frames.
4. Control Structures:
 - If statements.
 - For and while loops.
 - Using `apply()` functions.

Session 3: Data Manipulation and Visualization (2 hours)

1. Data Import and Export:

- Reading and writing data from/to various formats (CSV, Excel, etc.).

2. Data Cleaning and Transformation:

- Dealing with missing values.
- Removing duplicates.
- Transforming and recoding variables.

3. Basic Plotting:

- Creating scatter plots, bar plots, and line plots using base R.
- Adding titles, labels, and legends.

4. Introduction to ggplot2:

- Grammar of graphics principles.
- Creating more complex and customized plots.

Session 4: Functions and Packages (1 hour)

1. Creating Functions:

- Defining custom functions.
- Passing arguments and returning values.

2. Introduction to Packages:

- Installing and loading packages.
- Utilizing popular packages (dplyr, tidyr) for data manipulation.

3. Case Study: Data Analysis Workflow:

- Walking through a simple data analysis using R.

Module 03: Introduction to Descriptive Statistics (2 hours)

- Understanding the concepts of population and sample.
- Types of variables and scales of measurement.
- Measures of central tendency: mean, median, mode.
- Measures of dispersion: range, variance, standard deviation.

Module 04: Presentation of Data (2 hours)

- Organizing data with class limits, boundaries, midpoint, and interval.
- Creating frequency distributions and cumulative frequency distributions.
- Visualizing data using bar charts, pie charts, histograms, and frequency polygons.
- Exploring the 5-number summary and box plots.
- Creating scatter plots for bivariate data analysis.

Module 05: Random Variables and Probability (3 hours)

- Discrete and continuous random variables.
- Probability concepts and rules.
- Conditional probability and Bayes' theorem.
- Bernoulli, binomial, and Poisson distributions.
- Introduction to the normal distribution and standard scores.

Module 06: Correlation and Regression (3 hours)

- Exploring the correlation between two variables.
- Calculating and interpreting the correlation coefficient.
- Simple linear regression analysis.
- Interpreting regression coefficients and the coefficient of determination (r^2).

Module 07: Time Series Analysis (3 hours)

- Understanding time series data and its components.
- Decomposition of time series: trend, seasonality, and noise.
- Forecasting techniques: moving averages, exponential smoothing.
- Detecting and handling non-linear trends.

Module 08: Statistical Sampling (3 hours)

- Differentiating between sampling methods: random, systematic, stratified, cluster.
- Estimating sampling error and non-sampling error.
- Determining sample size using various methods.

Module 09: Hypothesis Testing and Inferential Statistics (4 hours)

- Formulating null and alternative hypotheses.
- Confidence intervals and level of significance.
- Understanding p-values and making decisions using them.
- Types of errors: Type I and Type II.
- Statistical power and sample size determination.

Module 10: Statistical Quality Control (3 hours)

- Introduction to quality control concepts.
- Causes of variation in processes.
- Control charts for monitoring process variables.
- Control charts for attribute data.
- Acceptance sampling and sampling plans.

Module 11: Experimental Design (3 hours)

- Introduction to experimental design principles.
- Randomization, replication, and control groups.
- Factorial designs and interactions.
- ANOVA (Analysis of Variance) for comparing means.

Module 12: Advanced Topics (7 hours)

- Multivariate analysis techniques: PCA (Principal Component Analysis), factor analysis.
- Bayesian statistics: Bayes' theorem, Bayesian inference.
- Integrating machine learning and statistics.
- Handling big data using statistical methods.
- Case studies and practical applications.

Mock Interview Session!

To Enroll in Course:

Contact: +8801704265972 (Call/WhatsApp)

[Sohan Khan](#), Course Coordinator at aiQuest Intelligence

Watch Free Courses: <https://www.aiquest.org/free-courses>

Facebook Community: [Join Our Community!](#)

Visit Our Pages: [Study Mart](#) , [aiQuest Intelligence](#)