



# Statistics for Data Science

[www.fractionshub.com](http://www.fractionshub.com)

# The WHY: Learning Objectives

- Statistics for Datascience aims to develop mathematical background to statistical methodologies extensively used in domains of 'Data-science'.
- The course helps you understand the driving screws of in-built R/Python functions.
- The course is tailored to suit students of diverse background. Building adequate theory, we explore routine and tricky questions, some of which appear in Datascience interviews.

# The WHO and WHEN:

## Instructors

Anees Parwez

---



Completed B.Stat  
from ISI Kolkata.  
Currently pursuing  
M.Stat at ISI Kolkata

and will be joining JP Morgan  
Chase as an Associate next year.

Souhardya Ray

---



Completed B.Stat  
from ISI Kolkata.  
Currently pursuing  
M.Stat at ISI Kolkata

and will be joining Morgan Stanley  
as an Associate next year.

## Class schedule

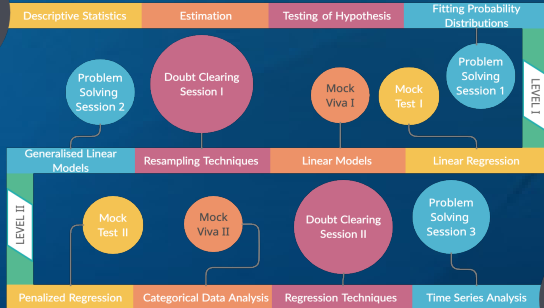
Monday: 9pm-11pm

Saturday: 9pm-11pm

# The WHAT: Syllabus Schematics

## STATISTICS FOR DATA SCIENCE : Course Plan & Description

Statistics



Miscellaneous Problems

# Lecture-Wise Syllabus

## Lectures (1-4)

---

- Descriptive Statistics.
- Estimation & Testing of Hypothesis.
- Linear Regression.
- Generalized Linear Models.

## Lectures (5-8)

---

- Categorical Data Analysis.
- Regression Techniques.
- Time Series Analysis.
- Miscellaneous Topics in Statistics.

## Brain Teasers and FAQ's

- Two Bi-variate datasets have positive correlation coefficients. Can the combined dataset have a negative correlation?
- Suppose there are  $n$  points on  $\mathbb{R}^2$ , show that a LAD line passes through at least 2 of the given  $n$  points.  
Note: LAD line means Least Absolute Deviation Minimising Line.
- Suppose  $X, Y$  are two Random Variables with correlation  $\rho$ . What is the correlation between  $X^2, Y^2$ ?

## Brain Teasers and FAQ's [continued]

- Suppose we have 5 independent Random Variables  $X_1, \dots, X_5$  such that  $\mathbb{E}[X_k] = k \cdot \mu$  for  $1 \leq k \leq 5$  and  $\text{Var}[X_k] = k$  for  $1 \leq k \leq 5$ . Find the Best Linear Unbiased Estimator of  $\mu$  ?  
Note: Here Best is in the sense of Minimized Mean Squared Error.
- Given the unit disk  $X^2 + Y^2 \leq 1$  simulate a uniformly random point inside it.  
Note: Uniformly Random on the Area.