Introduction to SDA

Dr. Mainak Thakur

10.08.2020

Syllabus

The following list of topics is tentative. Based on available time slots, some topics may be dropped or added or reordered. This course is divided into the following modules:

• Introduction to Statistical Thinking

• Review of Probability

• Overview of Exploratory Data Analysis

• Overview of Sampling Theory

• Important Univariate Distributions

• Estimation Theory

• Hypothesis Testing

• Multivariate Descriptive Statistics

Syllabus Contd.

- Multivariate Normal Distribution and properties

- Regression Analysis: Multiple Linear Regression

- Overview of Different Multivariate Techniques:

Cluster Analysis

Principal Component Analysis

Factor Analysis

- Time Series Analysis

- Application using Matlab/R/Python

Books

- **Applied Multivariate Statistical Analysis by Richard Johnson & Dean Wichern, Pearson**

- **An Introduction to Multivariate Statistical Analysis by T W Anderson, John Wiley & Sons Inc.**

- **Basic Econometrics by Damodar N. Gujarati, Tata McGraw-Hill**

- Applied Statistics and Probability for Engineers (3rd ed.) by Montgomery, D. C., 2002.

- Statistical Methods (Combined Volume) by N G Das, Tata McGraw-Hill.

- Time Series Analysis (4th ed.) by George E. P. Box, Gwilym M. Jenkins and Gregory C. Reinsel, John Wiley & Sons, 2013.
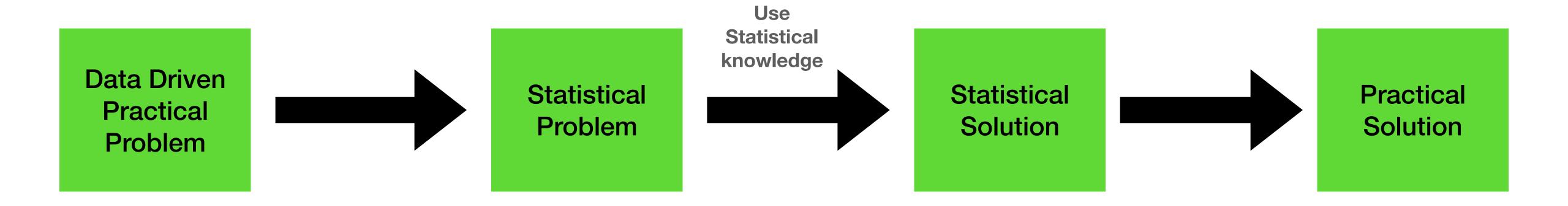
Evaluation Process (Tentative)

- Mid Sem 1: 10% (September)

- Mid Sem 2: 10% (November)

- Project: 30% (October/November)

- Quiz: 20% (Surprise, online test, class performance: through out the course)

- Assignment: 30% (through out the course)

Why study SDA?

- To introduce with the world of statistics

- Learn Statistics, fundamental for data science

- Analyse data and interpret knowledge out of it

- Apply in real life scenario

Practical problem & practical solution

| Data Driven Practical Problem | → | Statistical Problem | Use Statistical knowledge → | Statistical Solution | → | Practical Solution |

- An example to understand

Thank You

Any Questions?