Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach to analyzing data to summarize their main characteristics, often with visual methods.

In this exercise following points are to be discussed:

1. Data and types of data e.g. continuous or discrete.
2. Distribution: Some Common Continuous (Normal, Exponential, Uniform) and Discrete (Binomial, Poisson) and Concept of Parameter
3. Exploratory Data Analysis (EDA)
    - EDA for Individual Variables:
      Summarization: Measures of Central Tendency, Measures of Dispersion.
      Data Visualization: Histogram/Bar Chart, Box Plot, Stem and Leaf Display.
      Outlier Detection, Symmetry Checking.
      Testing for Normality: Histogram, Quantile-Quantile Plot, Kolmogorov-Smirnov Test, Shapiro-Wilk Test.

    - EDA for Multiple Variables:
      Pair wise Scatter Plots
      Correlation Analysis

Case Study

Consider the data set *swiss* and perform the following:

(i) What does this data set represent? What is the source of data set?

(ii) Discuss each variable and its data type.

(iii) Perform exploratory data analysis on this data set.
For each variable: (a) Summarization, (b) Visualization: Histogram, Stem and Leaf Plot, Q-Q plot. Normality Test: K-S Test, Shapiro-Wilk Test.

(iv) Regression Analysis: Perform regression analysis of Fertility and Agriculture variables.