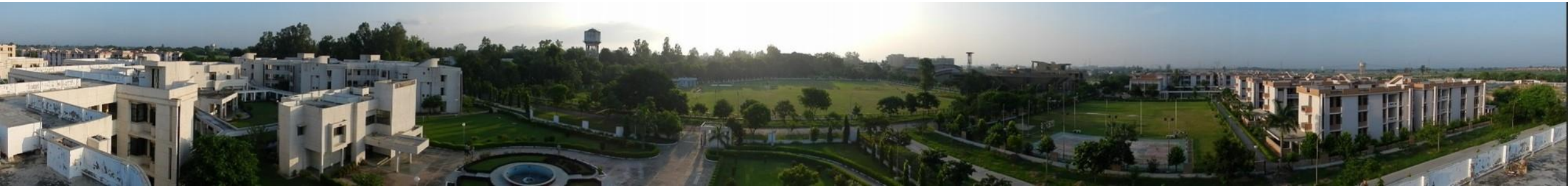
Fundamentals of R



Introduction

- Introduction of R and How to download and install it
- Starting to working with R
- Basic operations with R
- Datatypes in R
- User Defined Functions with R

Introduction to R

- R is an independent, open-source tool
- R is a statistical computing environment that includes an interpreter for the R programming language
- Download R: https://cran.r-project.org/bin/windows/base/
- Download R Studio: https://www.rstudio.com/products/rstudio/download/#download
- We will discuss and understand various functionalities of R assuming that one does not have a coding background

Thanks!

INDIAN INSTIUTE OF TECHNOLOGY KANPUR

Case Study: Auto parts manufacturing company Xio Ltd

BACKGROUND

- An auto parts manufacturing company Xio Ltd. conducts regular training and development programs for its 1000 employees.
- It wants to check the current level of satisfaction of employees and that after the training program.
- The firm has given you the primary survey data response feedback taken from these 1000 employees pre- and post-training program and wants you to examine the impact of the training program on the employee satisfaction

DATA

- The data file ("Data.csv") pertaining to this case study comprises 3 variables: (1) Employee ID, (2) Pre T&D survey, (3) Post T&D survey
- Pre column represents survey responses on a scale of 1-8 before the training and development program
- Post column represents survey responses on a scale of 1-8 after the training and development program
- Often individuals are biased on upside and downside, slight correction is made to scale this bias
- This bias correction converts the original data (responses in the integer form) into fractional form

Case Analysis

Summarize and visualize the survey response data to conduct a preliminary analysis of the data (using R tool)

- Provide a broad summary of the data, its structure and overview
- Visualize the data and contrast the key variables
- Examine the measures of central tendency: Mean, Media, and Mode
- Examine the measures of variability: Range, Variance, Standard Deviation, and Mean Absolute Deviation
- Examine the shape parameters: Skewness, Kurtosis, and Normality of the data

Case Analysis

Using the probability sampling techniques, carry out the following analysis (using R tool)

- Explain and implement the following approaches
  - Simple random sampling
  - Systematic Sampling
  - Stratified Sampling
  - Cluster Sampling
- Summarize and visualize the samples Pre and Post training and development program
- Compute the measures of central tendency, range, and dispersion for the Pre and Post samples
- Compare the sample Pre and Post distributions with each other and normal distributions

Case Analysis

Using the interval estimation techniques, carry out the following analysis (using R tool)

- Explain and implement confidence interval estimation for our Pre and Post survey response data
  - Using Normal distribution
  - Using Student's t-distribution
- Compare the results from Normal distribution and t-distribution confidence intervals
- Visualize the differences between Normal distribution and t-distribution for different sample sizes
- Model the 'Detractor (0)' and 'Promoter (1)' data using binomial distribution, and make inferences about the population parameter and employee-satisfaction

Recap of Xio Ltd. Case Study

Let us recap the Xio ltd. case study problem

- An auto parts manufacturing company Xio Ltd. conducts regular training and development programs for its 1000 employees
- You have already analyzed the data with the following tools
  - You have summarized and visualized the data
  - You have computed various inferential statistics (measures of central tendency, variation, shape, etc.)
  - You have taken various probability samples to make inferences about the Pre and Post training population
  - You have performed confidence interval estimation using Normal, t, and Binomial distributions

Case analysis

Using the hypothesis testing you are expected to carry out the following analysis (using R tool)

- Design single sample hypothesis tests for Pre and Post sample data
  - Conduct hypothesis testing with z-statistics
  - Conduct hypothesis testing with t-statistics
- Design two sample hypothesis tests for Pre and Post sample data
  - Conduct hypothesis testing with z-statistics
  - Conduct hypothesis testing with t-statistics

Case analysis

Using the hypothesis testing you are expected to carry out the following analysis (using R tool)

- Convert the survey response data into binary (0,1) form and define employees as detractors (satisfaction score of less than 4.5) and promoters (satisfaction score of more than 4.5)
- Conduct one sample and two sample tests of population proportions using binomial distribution
- Examine the Pre and Post sample data and make inferences about population parameters
- Examine, whether the Pre sample population and Post sample population has different proportion of detractors and promoters
- Examine, if there is indeed some impact of training and development in increasing the proportion of promoters, and whether it is statistically significant

Thanks!