Summarizing Data

Md. Aminul Islam Shazid

Outline

1. Introduction
2. Tabular Summary

Introduction

Summarizing Data

- Raw data by itself is not useful
- Need to extract insight from data
- Data needs to be summarized to gain insights
- Can summarize in two ways:
  - Tabular summary
  - Graphical summary

Tabular Summary

Tabular Summary

Frequency Table

Frequency Table

- Organizes data from the sample by listing distinct values or classes
- Shows the frequency of each class in the sample
- Can be constructed for categorical or numerical data:
  - For categorical variable, it shows the number of observations in each category
  - For discrete numeric variable, it shows how many times each value has been observed
  - For continuous numeric variable, classes or bins are formed first, then the table shows how many values fall in each class
- Forms the basis for graphical summaries like bar charts and histograms

Example: Frequency Table for Categorical Variable

Suppose we have a discreete variable with four categories: A, B, C and D. The data in the sample is: A, B, A, D, A, A, B, C, C, B, A, A, C, D, D, B, D, C, C, B, C. The Frequency table would be:

Category	Tally	Frequency
A	𝍸	6
B	𝍸	5
C	𝍸	6
D	𝍸	4

The tally column is only used to keep track of the data when counting by hand.

One can add a percentage column in necessary.

Questions?