

# BASIC STATISTICS

DEEPAN PITCHAIRAJ

M.Sc Biostatistics

National Institute of Mental Health And Neurosciences(NIMHANS)  
Bengaluru, INDIA

29/06/2024



**NELIREF**

Data Science and AI

# Symposium

2024



# Table of Contents

## 1 Introduction

### ► Introduction

Data

Steps in Data Analysis

Types of Analysis

### ► Definition of Statistics

### ► Collection of Data

Population & Sample

Descriptive & Inferential Statistics

### ► Organizing Data

Qualitative

Quantitative

### ► Summarizing Data

# Data

## 1 Introduction

What, Where, Why, When and How?

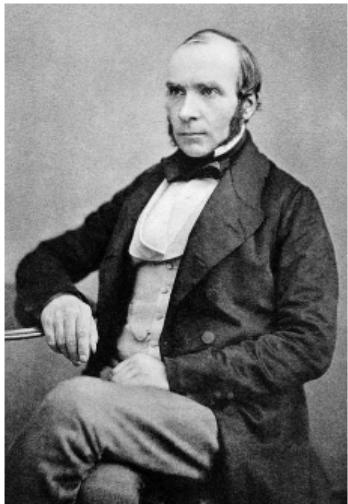
### Definition

Data is a collection of facts, information, and statistics and this can be in various forms such as numbers, text, sound, images, or any other format.

- Medicine and Healthcare
- Agriculture
- Manufacturing
- Finance and Banking
- and so on..

# Data

## 1 Introduction



*John Snow*

Discovery of Cholera



Governments & Business



YOU!

# Data

## 1 Introduction



Ensure data collection occurs during the event or select data from the specific past timeframe.

- **Primary Data**
  - Survey data
  - Experimental data
  - Observational data
- **Secondary Data**
  - Public Data
  - Private Data
  - Personal Data

<https://researchmethod.net/secondary-data/>

<https://careerfoundry.com/en/blog/data-analytics/web-scraping-guide/>



Scrapy



BeautifulSoup

Selenium

# Steps in Data Analysis

## 1 Introduction

- 1. Define the Problem:** Clearly outline the problem, expected deliverables, form a suitable team, and create an execution plan.
- 2. Prepare the Data:** Collect, characterize, clean, transform, and partition the data for further processing.
- 3. Implement the Analysis:** Select and optimize appropriate analysis techniques based on the problem definition and prepared data.
- 4. Deploy the Results:** Communicate and integrate the analysis results into existing processes.

Source: "*Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*"  
by Glenn J. Myatt (2007)

# Types of Analysis

## 1 Introduction

### Step 3: Implement the Analysis



#### Descriptive

What have we done in the past?



#### Diagnostic

Why have we seen past results?



#### Predictive

Where are we going and when?



#### Prescriptive

How should we take action?

# Table of Contents

## 2 Definition of Statistics

### ► Introduction

Data

Steps in Data Analysis

Types of Analysis

### ► Definition of Statistics

### ► Collection of Data

Population & Sample

Descriptive & Inferential Statistics

### ► Organizing Data

Qualitative

Quantitative

### ► Summarizing Data

# Meaning of Statistics

## 2 Definition of Statistics

The word statistics is normally referred either as numerical facts or methods.

**Singular Sense:** Statistics is concerned with scientific method for collecting, organizing, summarizing, presenting, analyzing and interpreting of data.

**Plural Sense:** In the plural sense, it refers to the numerical facts and figures systematically collected for some special purpose. This definition has the following features



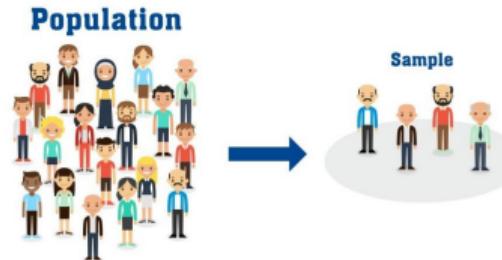
# Table of Contents

## 3 Collection of Data

- ▶ Introduction
  - Data
  - Steps in Data Analysis
  - Types of Analysis
- ▶ Definition of Statistics
- ▶ Collection of Data
  - Population & Sample
  - Descriptive & Inferential Statistics
- ▶ Organizing Data
  - Qualitative
  - Quantitative
- ▶ Summarizing Data

# Population & Sample

## 3 Collection of Data



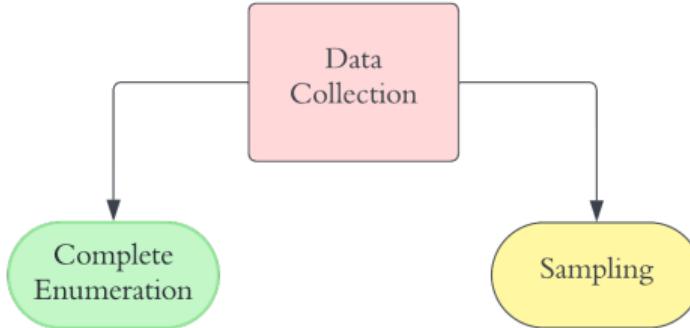
**Warning:** Don't expect data to be collected from Humans always.

**Population:** The word population or statistical population is used for aggregate of all units or objects within the purview of enquiry.

**Sample:** is small proportion of the population taken from the population to study the characteristics of the population.

# Descriptive & Inferential Statistics

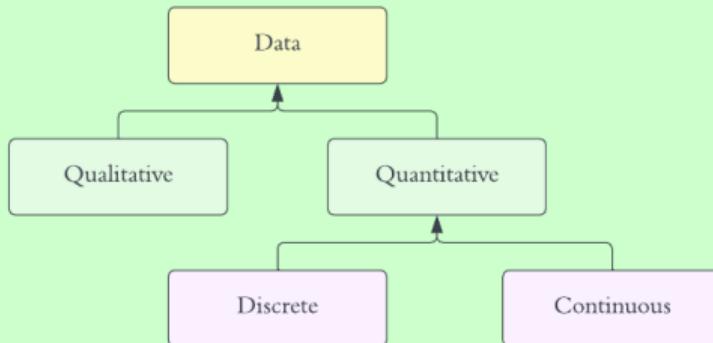
## 3 Collection of Data



- **Descriptive statistics** just describes the data in a condensed form for solving some limited problems. They do not involve beyond the data at hand.
- **Inferential statistics** are used when we want to draw meaningful conclusions based on sample data drawn from a large population.

# Types of Data based on its Nature

Note



## Scales of Measurement

- **Nominal Scale:** Categorical data where variables are named or labeled with no inherent order.
- **Ordinal Scale:** Data that can be ranked or ordered but the differences between values are not consistent.
- **Interval Scale:** Data where differences between values are consistent, but there's no true zero point.
- **Ratio Scale:** Data with consistent differences and a true zero point, allowing for ratios to be meaningful.

# Table of Contents

## 4 Organizing Data

### ► Introduction

Data

Steps in Data Analysis

Types of Analysis

### ► Definition of Statistics

### ► Collection of Data

Population & Sample

Descriptive & Inferential Statistics

### ► Organizing Data

Qualitative

Quantitative

### ► Summarizing Data

# Organizing Data:Qualitative

## 4 Organizing Data

Professor Weiss asked his introductory statistics students to state their political party affiliations as Democratic, Republican, or Other. The responses of the 40 students in the class are given in Table

### Sample Data

Democratic	Other	Democratic	Other	Democratic
Republican	Republican	Other	Other	Republican
Republican	Republican	Republican	Democratic	Republican
Republican	Democratic	Democratic	Other	Republican
Democratic	Democratic	Republican	Democratic	Democratic
Republican	Republican	Other	Other	Democratic
Republican	Democratic	Republican	Other	Other
Republican	Republican	Republican	Democratic	Republican

### Frequency distribution

Party	Tally	Frequency
Democratic		13
Republican		18
Other		9
		40

Source: "*Introductory Statistics*" by Neil A. Weiss (2017)

# Organizing Data: Quantitative

## 4 Organizing Data

### Sample Data

8	6	7	7	4	3	1	7
5	6	6	5	6	8	5	5
7	4	6	6	5	5	5	4

55.0	50.8	52.1	57.0	52.1	53.5
51.3	51.5	46.4	55.3	45.5	54.1
55.3	50.3	47.2	53.8	50.7	51.5
50.5	51.8	53.6	52.0	51.9	54.3
48.0	53.3	53.5	56.0	49.1	53.9

### Frequency distribution

#### Discrete

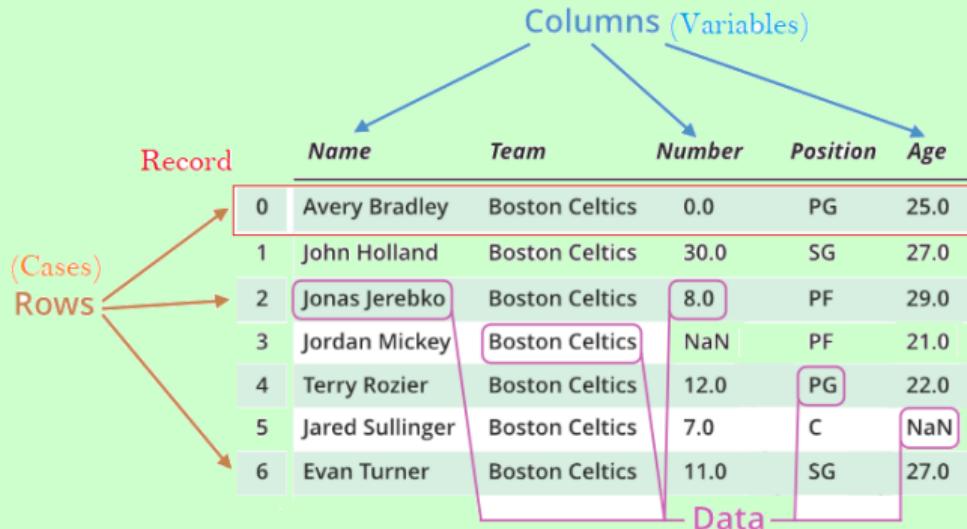
No of Family Members	Frequency
1	1
3	1
4	3
5	7
6	6
7	4
8	2
<b>Grand Total</b>	<b>24</b>

#### Continuous

Class Interval	Frequency
45.5 - 47.5	3
48.5 - 50.5	5
51.5 - 53.5	14
54.5 - 56.5	6
57.5 - 59.5	0
<b>Total:</b>	<b>28</b>

# Dataframe

Note



**Cases** are the individuals from which data are collected.

A **variable** is a characteristic that is measured and can take on different values. In other words, something that varies between cases.

Python Libraries



# Table of Contents

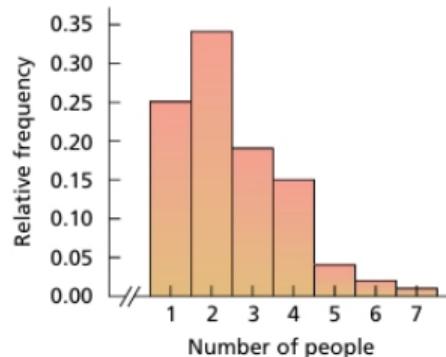
## 5 Summarizing Data

- ▶ Introduction
  - Data
  - Steps in Data Analysis
  - Types of Analysis
- ▶ Definition of Statistics
- ▶ Collection of Data
  - Population & Sample
  - Descriptive & Inferential Statistics
- ▶ Organizing Data
  - Qualitative
  - Quantitative
- ▶ Summarizing Data

# Summarizing Data

## 5 Summarizing Data

"Comprehension is compression"



	name	degree	score
0	aparna	MBA	90
1	pankaj	BCA	40
2	sudhir	M.Tech	80
3	Geeku	MBA	98

Measures of Central tendency, Dispersion & Shape

# Get your feet wet!

## 5 Summarizing Data

Let's dive into Google colab



(To be continued) Summarizing, Presenting, Analyzing of Data..