Stats for Data Analysis

Jubayer Hossain

Health Data Research Organization

September 15, 2020

Steps in Data Research

- Planning / Study Design
- Data Collection
- Data Preparation
- Data Analysis
- Interpretation
- Report Writing / Presentation

What is Data?

- Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things.
- Data - tool to reach suitable conclusion.

Sources of Data

- Primary sources of data
- Secondary sources of data

Primary Sources of Data

- Collection of data from source of origin
- Conducting interviews, experimentation
- Provide first hand information

Secondary Sources of Data

- Collection of data from agency which already has collected data and processed it.
- Conducting interviews, experimentation
- Provide first hand information

Types of Data

- Primary Data
  - Data collected from primary sources
  - Personal interviews, Mailing questionnaire, Telephone interviews
  - Collected first time
- Secondary Data
  - Data collected from primary sources.
  - Published sources – Govt. publications, private publications, reports of NGOs

Important Agencies of Secondary Data

- https://dghs.gov.bd/index.php/en/data
- https://framinghamheartstudy.org/
- https://www.data.gov/
- https://healthdata.gov/
- https://www.who.int/ictrp/network/trds/en/
- https://data.cdc.gov/
- https://www.kaggle.com/

Terms used in Data Collection

- **Variable** – values which changes
- **Observations** – values from variables are referred as observations.
- **Statistical Investigation** – search for information conducted by using statistical methods.
- **Investigator** – who conducts statistical inquiry.
- **Enumerators** – person hired by investigator to collect data.
- **Respondents** – from whom information is collected.

Data Collection Methods

- Interviews
- Questionnaires and surveys
- Observations
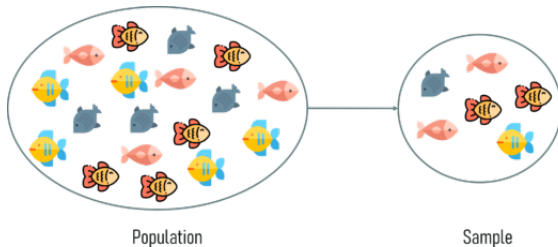- Documents and records
- Focus groups
- Oral histories
  See More https://www.jotform.com/data-collection-methods/

What is Statistics?

- Statistics is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data.
- In applying statistics to a scientific, industrial, or social problem, it is conventional to begin with a statistical population or a statistical model to be studied.

Basic Terminologies In Statistics

- **Population:** A collection or set of individuals or objects or events whose properties are to be analyzed
- **Sample:** A subset of the population is called 'Sample'. A well-chosen sample will contain most of the information about a particular population parameter.

Population   Sample

https://www.edureka.co/blog/statistics-and-probability/

Types of Statistics

- Descriptive Statistics
- Inferential Statistics

Descriptive Statistics

Descriptive statistical methods provide an exploratory assessment of the data from a study

- Exploratory data analysis techniques
- Organization and summarization of data
    - Tables
    - Graphs
    - Summary measures

Inferential Statistics

Assess the strength of evidence for/against a hypothesis; evaluate the data

- Inferential statistical methods provide a confirmatory data analysis
  - Generalize conclusions from data from part of a group (sample) to the whole group (population)
  - Assess the strength of the evidence
  - Make comparisons
  - Make predictions

Types of Variables

- Qualitative / Categorical
  - Classifies individuals or items into different groups
- Quantitative / Numerical
  - Numerical, measurable quantities in which arithmetic operations often make sense
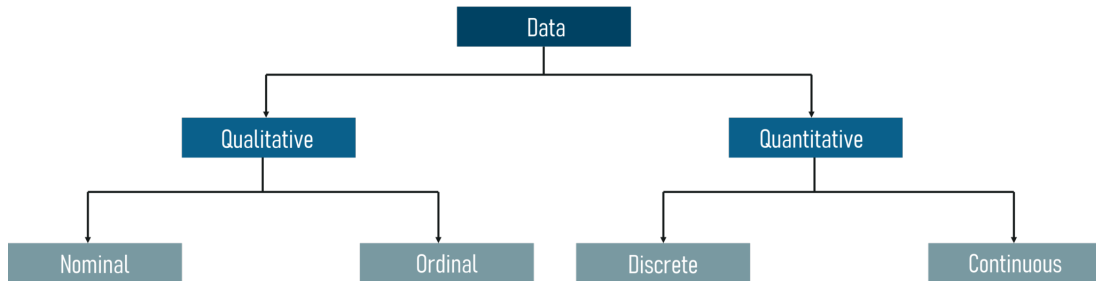
Types of Variables Cont ...

Image Sources: https://www.edureka.co/blog/statistics-and-probability/

Types of Variables Cont . . .

- **Continuous** – could take on any value within an interval, many possible values
- **Discrete** – countable value, finite number of values
- **Ordinal** – groups have an order or ranking
- **Nominal** – groups are merely names, no ranking

Dataset

	A	B	C	D	E	F	G
1	RANDID	SEX	AGE	SYSBP	DIABP	TOTCHOL	
2	9651782	2	56	134	75	217	
3	9309361	1	63	110	68	217	
4	8224634	2	54	135	85	230	
5	6215341	1	44	134	88	240	
6	2930353	2	50	156	90	298	
7	5251766	1	52	125	82	219	
8	602184	2	40	124	76	197	
9	6611751	1	44	138	98	230	
10	7983169	2	44	113	78	178	

THE END