Lecture No. 1

Introduction to Statistics
Statistics and Probability

Dr. Shabbir Ahmad

Assistant Professor,
Department of Mathematics,
COMSATS University
Islamabad, Wah Campus

What is Statistics, Types of Statistics, Types of Data and Sources, Scales of Measurement

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

In this lecture

- What is Statistics?
- Data, Data Sets, Elements, Variables and Observations
- Population vs. Sample
- Descriptive and Inferential Statistics
- Data Sources (Primary and Secondary Data)
- Types of Data (Qualitative and Quantitative)
- Scales of Measurement: Qualitative (Nominal, Ordinal) and Quantitative (Ratio, Interval)
- Cross-Sectional Data, Time Series Data

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

What is Statistics?

- The Science of extracting information from data.
- Statistics is the methodology which statisticians have developed for interpreting and drawing conclusions from data.

- Statistics involves

1. Identifying the problem.
2. Collecting the data.
3. Organizing or manipulating data.
4. Presenting and
5. summarizing the data.
6. Analyzing the data.
7. Drawing conclusion.
8. Making inferences.

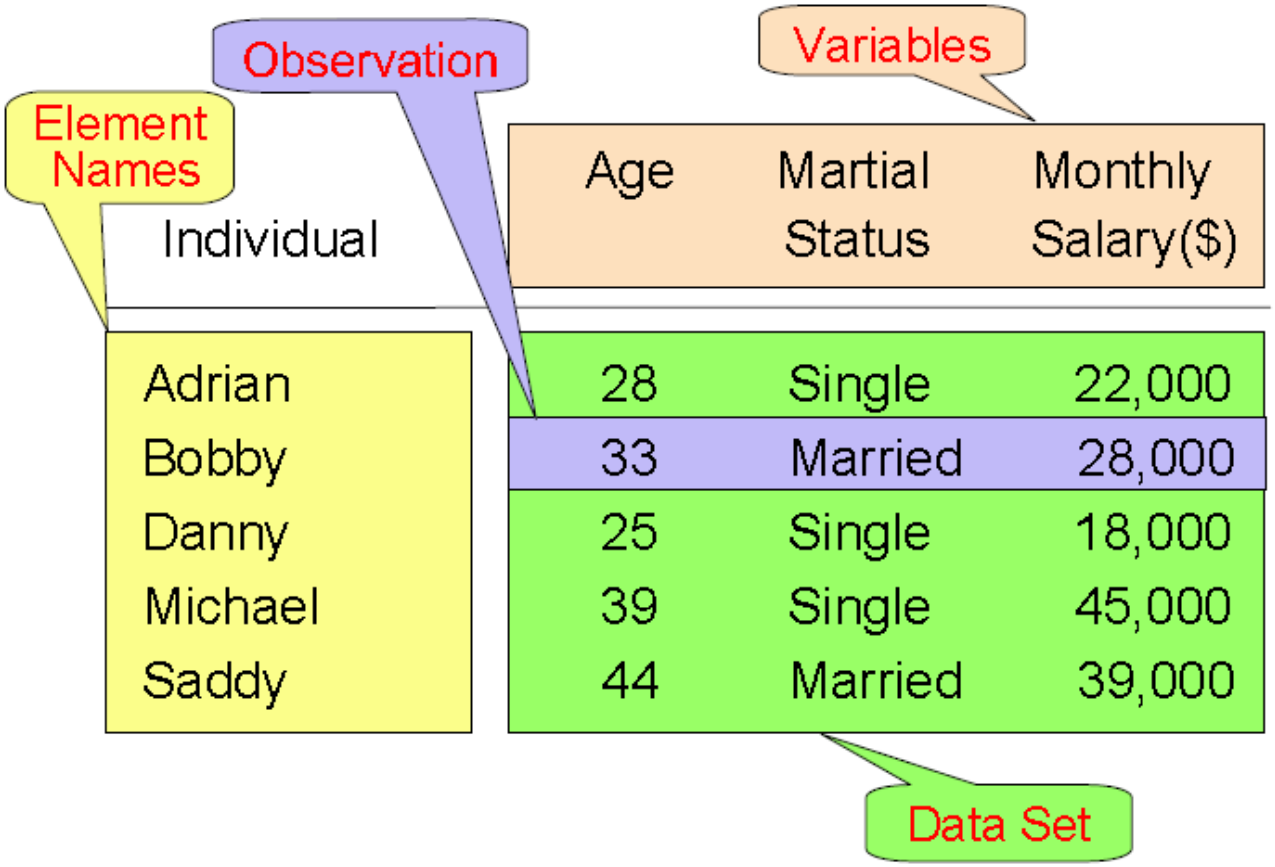
Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:26:25 AM

Data, Data Sets, Elements, Variables and Observations

- **Data** are the facts and figures collected, summarized, analyzed, and interpreted.
- The data collected in a particular study are referred to as the **data set**.
- The **elements** are the entities on which data are collected.
- A **variable** is a characteristic of interest for the elements.
- The set of measurements collected for a particular element is called an **observation**.
- The **total number of data** values in a data set is the number of elements multiplied by the number of variables.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Data, Data Sets, Elements, Variables and Observations



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Key Definitions

- A **population** is the collection of all members of a group
- A **sample** is a portion of the population selected for analysis
- A **parameter** is a numerical measure that describes a characteristic of a population
- A **statistic** is a numerical measure that describes a characteristic of a sample

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Population vs. Sample

Population

- Census survey: Collect data in a given population.
- Numerical measures calculated by using all the data in a given population are called parameters.

Sample

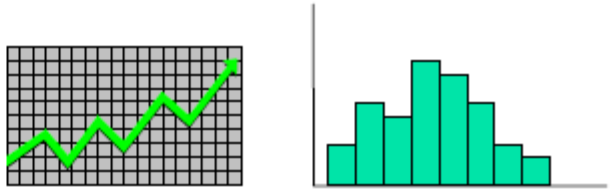
- Sample survey: Collect data for a sample.
- Numerical measures computed from sample data are called statistics.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Descriptive Statistics

Descriptive statistics are the tabular, graphical, and numerical methods used to summarize data.

- **Present data**
  - e.g., Tables and graphs

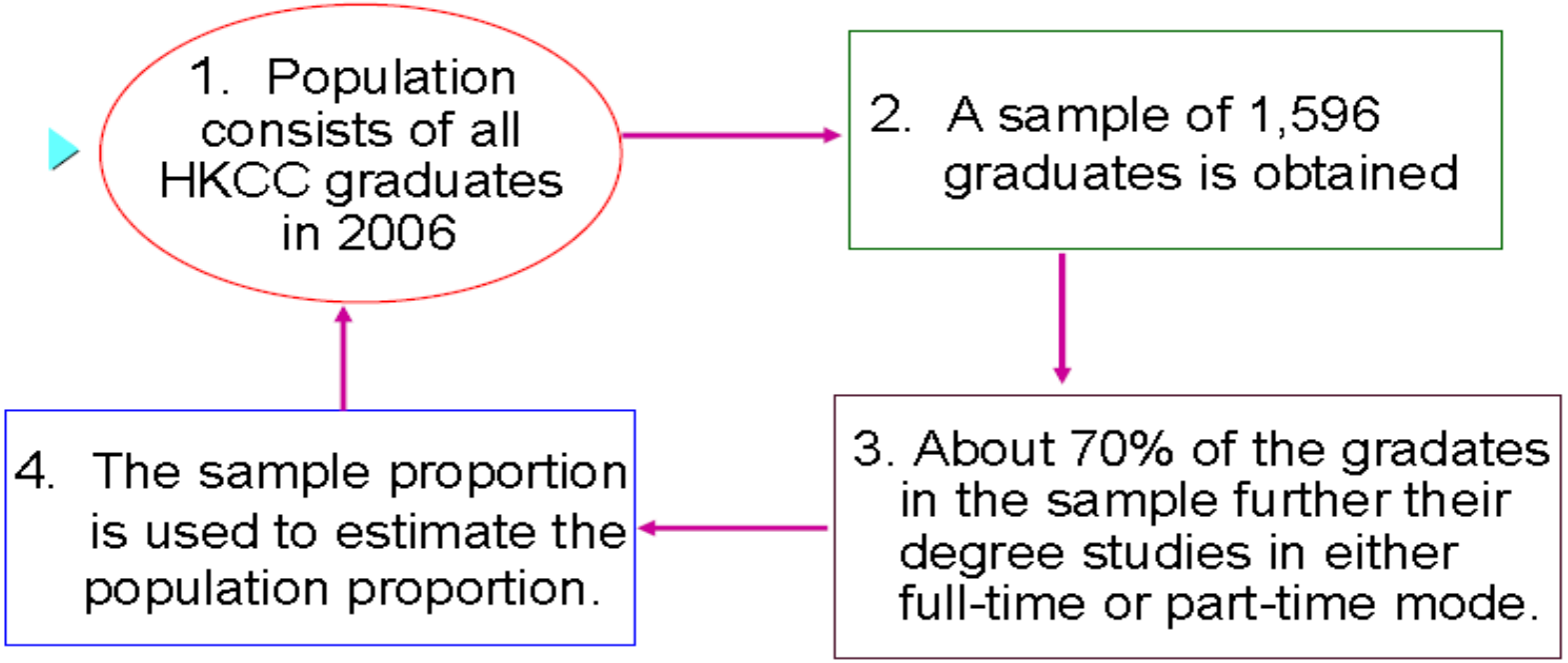


- **Characterize data**
  - e.g., Sample mean = $\dfrac{\text{Sum of data}}{\text{Sample size}} = \dfrac{\sum x_i}{n}$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Inferential Statistics

The process of using data obtained from a sample to make estimates or test hypotheses about the characteristics of a population.



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Inferential Statistics

- **Estimation**

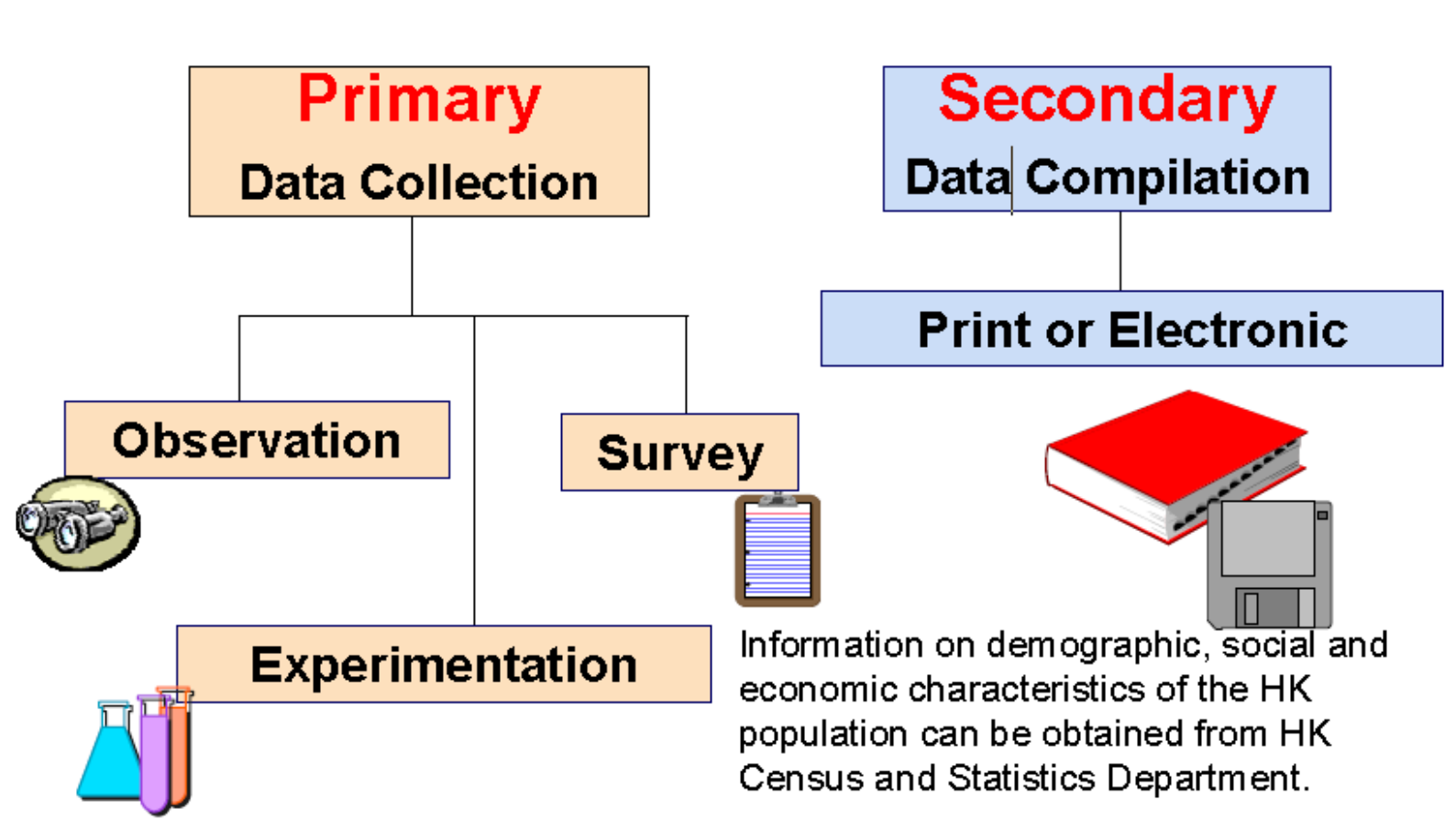
e.g. Estimate the population mean cholesterol level using the sample mean cholesterol level

- **Hypothesis testing**

e.g. Do the sample data show evidence that the population mean cholesterol level is above 5 mg?

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Data Sources



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Data Sources

Internal sources . E.g. Data available internally at HKCC.

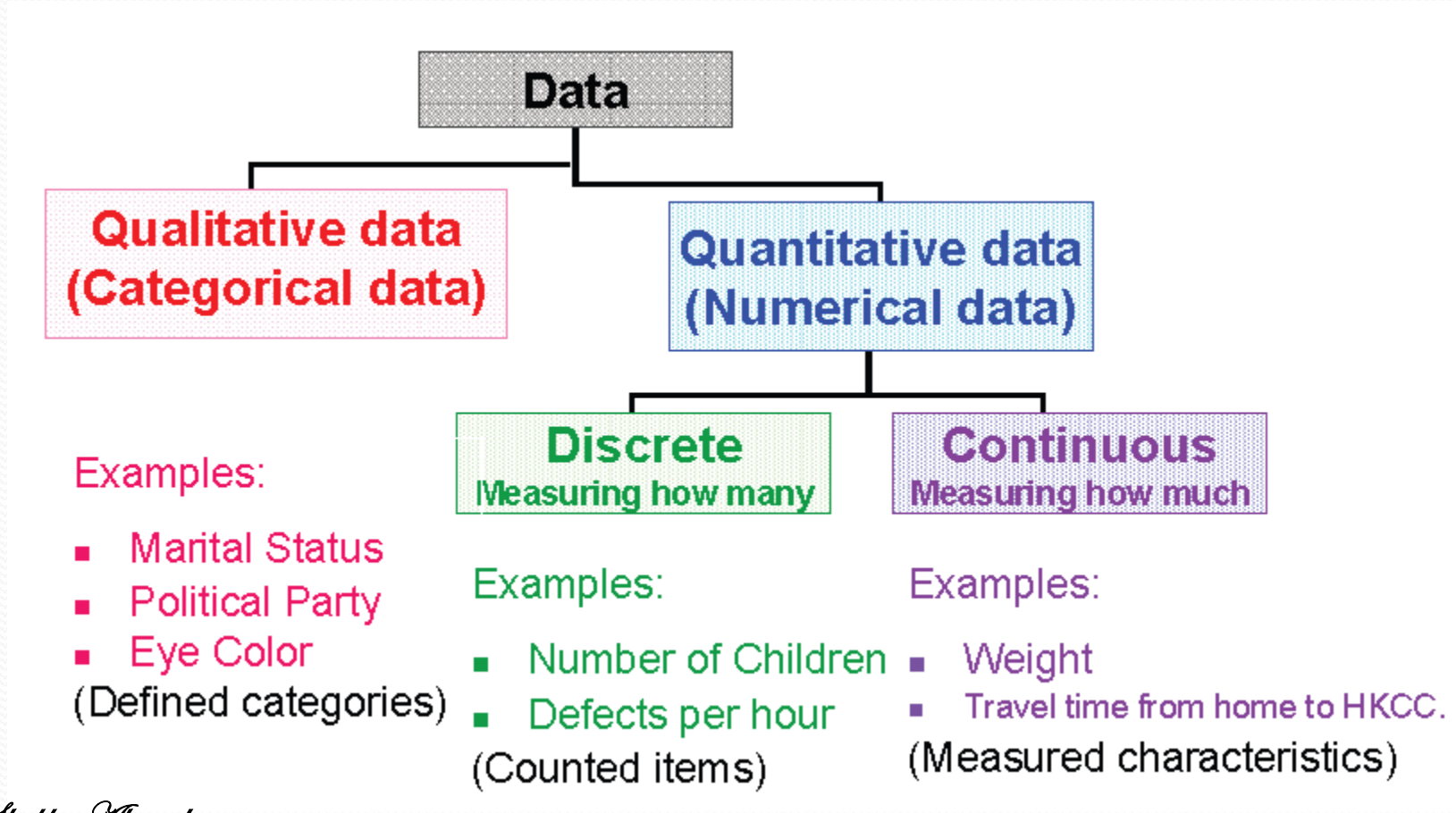
External sources . E.g. Information on demographic, social and economic characteristics of the HK population can be obtained from HK Census and Statistics

Experiment: One or more variables are identified and controlled so that data can be obtained about how they influence the variable of interest. E.g. Test the effectiveness of a new drug.

Observational study: The study make no attempt to control the variable of interest. E.g. Sample survey or census survey.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

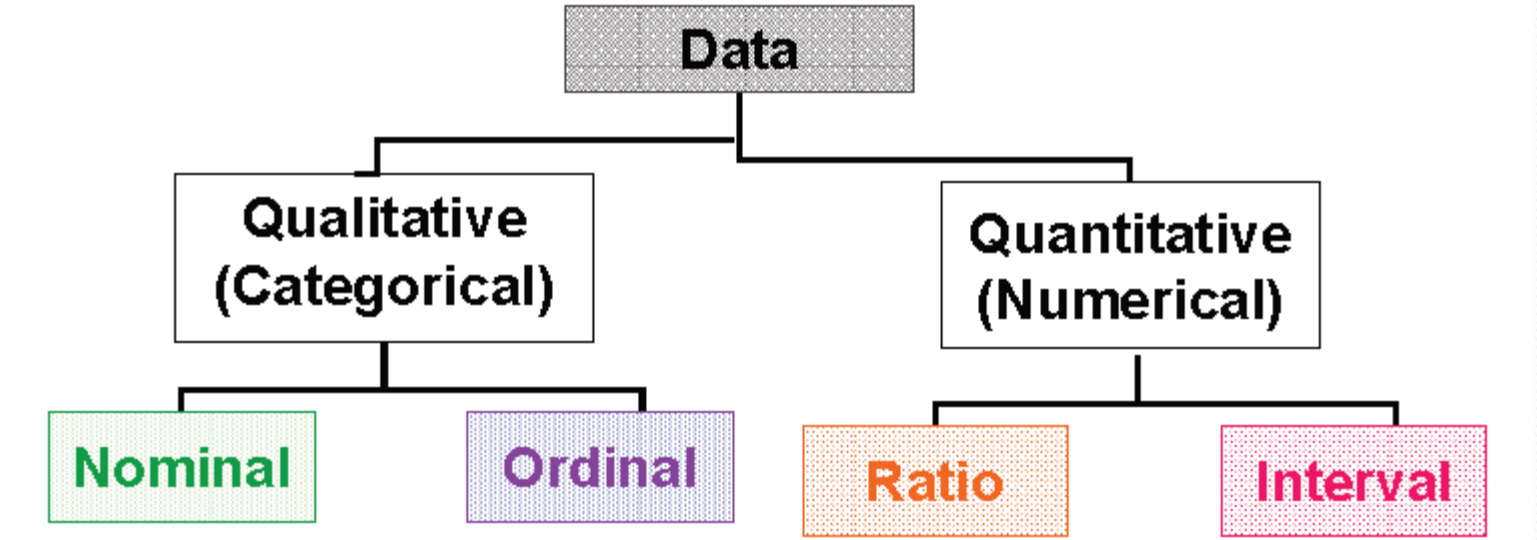
Types of Data



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Scales of Measurement

We have previously classified data into two types: Qualitative data and Quantitative data. Regarding the scales of measurement, we have: nominal level data, ordinal level data, ratio level data, and interval level data.



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Scales of Measurement

QUALITATIVE or CATEGORICAL DATA

The data values are categories which represent group membership. This kind of measurement usually involves naming or labeling.

- **NOMINAL LEVEL DATA:**

Two values cannot be compared mathematically.eg. marital status, blood type, gender, type of cars

- **ORDINAL LEVEL DATA:**

The values can be ordered.eg. teaching ability, grades, service quality rating,

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Scales of Measurement

QUANTITATIVE or NUMERICAL DATA

Data that consists of numerical measurement are called quantitative or numerical data. The data values are numbers that represent quantity.

- **RATIO LEVEL DATA:**

The data can be ordered. Differences make sense. The ratio of two data also has a meaningful interpretation. This scale requires that a zero value be included to indicate that nothing exist for the variable at the zero point.eg. height, weight, age, length, weekly food spending

- **INTERVAL LEVEL DATA:**

The data can be ordered. Differences make sense, but ratios do not. E.g. Temperature in Fahrenheit, or in Celsius, dates.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Types of Data

1. Age: _____

2. Sex: _____

3.How many mobile phones have you bought during the last 2 years?

4.Are you satisfy with your current mobile phone service provider?

___unsatisfied___satisfied

___very satisfied

Q: Which variables are qualitative and which variables are quantitative?

Q: What type of measurement scale is used for each of the variables?

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Cross-Sectional Data

Cross-sectional data are collected at the same or approximately the same point in time.

A cross-sectional study collects data on study units at some fixed time. Different subjects are usually compared to each other at one point in time.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Time Series Data

Time series data are collected over several time periods.

A longitudinal study collects information on study units over a specified time interval. Subjects are followed over time, and compared among themselves at different point in time.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Question 1

A Japanese motor company has been advertising a series of comparisons between its cars and American's car. The following variables were obtained from the study. For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the level of measurement.

(i) Whether an air conditioner is standard equipment on a car.

(ii) Driver's rating of the handling characteristics.

(iii) Sound level measured in decibels inside the car.

(iv) Kilometer per gallon for the cars.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Question 2

- For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the level of measurement.

(i)Average height of Hong Kong primary school students.

(ii) Number of mobile phones you have.

(iii) Annual salary income of a policeman.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Question 3

- A manufacturer of a Japan car is planning to survey car owners in Hong Kong to determine their purchasing habits. Among the questions to be included are those that relate to

(1) whether their current car is second-hand or not;

(2) the number of car do they own.

(i) Describe the population and frame.

(ii)Develop two categorical or numerical questions that would be appropriate to this survey.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM

Answers

Question 1

(i) Categorical, nominal **(ii)** Categorical, ordinal **(iii)** Numerical, continuous, ratio **(iv)** Numerical, continuous, ratio

Question 2

(i) numerical, continuous, ratio scale **(ii)** numerical, discrete, ratio scale **(iii)** numerical, continuous, ratio scale

Question 3

(i) Population : All Car owners in Hong Kong

(ii) Sample frame : Households in Hong Kong

Question 3

(iii)(1) Whether your current car is a second-hand or not?

(2) How many cars do you have?

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 11:24:55 AM