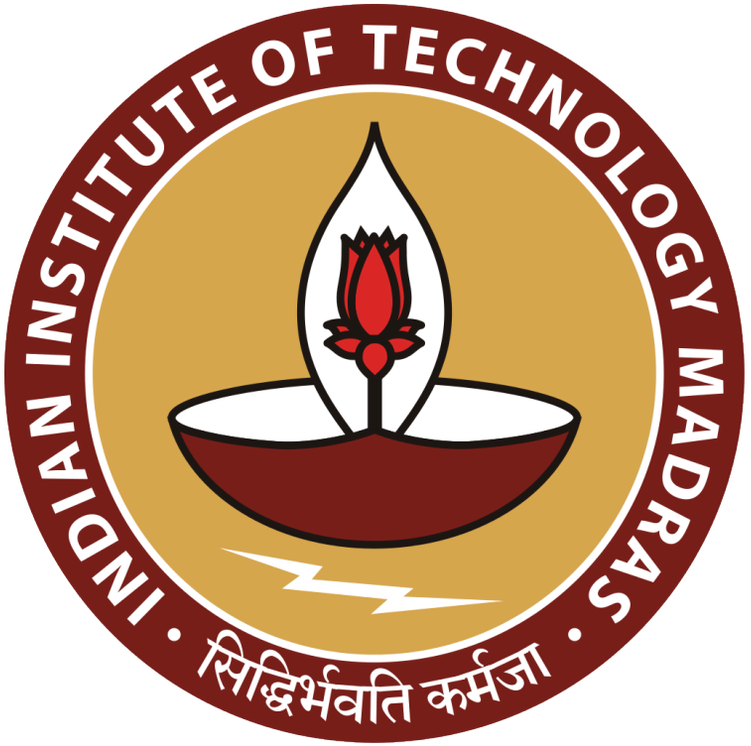


IIT Madras

ONLINE DEGREE

Statistics for Data Science -1

Introduction and types of data

Usha Mohan

Indian Institute of Technology Madras

Learning objectives

1. What is statistics?
   - Descriptive statistics, inferential statistics.
   - Distinguish between a sample and a population.
2. Understand how data are collected.
   - Identify variables and cases (observations) in a data set
3. Types of data-
   - classify data as categorical(qualitative) or numerical(quantitative) data.
   - Understand cross-sectional versus time-series data.
   - Measurement scales
4. Creating data sets; Downloading and manipulating data sets; working on subsets of data.
5. Framing questions that can be answered from data.

Introduction

- Basic definitions
- Population and sample

Understanding data

Classification of data

- Categorical and numerical
- Cross-sectional versus time-series data
- Scales of measurement

Scales of measurement

- Data collection requires one of the following scales of measurement: nominal, ordinal, interval, or ratio.

Nominal scale of measurement

- When the data for a variable consist of labels or names used to identify the characteristic of an observation, the scale of measurement is considered a nominal scale.

Nominal scale of measurement

- When the data for a variable consist of labels or names used to identify the characteristic of an observation, the scale of measurement is considered a nominal scale.
- Examples: Name, Board, Gender, Blood group etc.

Nominal scale of measurement

- When the data for a variable consist of labels or names used to identify the characteristic of an observation, the scale of measurement is considered a nominal scale.
- Examples: Name, Board, Gender, Blood group etc.
- Sometimes nominal variables might be numerically coded.

Nominal scale of measurement

- When the data for a variable consist of labels or names used to identify the characteristic of an observation, the scale of measurement is considered a nominal scale.
- Examples: Name, Board, Gender, Blood group etc.
- Sometimes nominal variables might be numerically coded.
  - For example: We might code Men as 1 and Women as 2. Or Code Men as 3 and Women as 1. Both codes are valid.

Nominal scale of measurement

- When the data for a variable consist of labels or names used to identify the characteristic of an observation, the scale of measurement is considered a nominal scale.
- Examples: Name, Board, Gender, Blood group etc.
- Sometimes nominal variables might be numerically coded.
  - For example: We might code Men as 1 and Women as 2. Or Code Men as 3 and Women as 1. Both codes are valid.
- There is no ordering in the variable.

Nominal scale of measurement

- When the data for a variable consist of labels or names used to identify the characteristic of an observation, the scale of measurement is considered a nominal scale.
- Examples: Name, Board, Gender, Blood group etc.
- Sometimes nominal variables might be numerically coded.
  - For example: We might code Men as 1 and Women as 2. Or Code Men as 3 and Women as 1. Both codes are valid.
- There is no ordering in the variable.
- Nominal: name categories without implying order

Ordinal scale of measurement

- Data exhibits properties of nominal data and the order or rank of data is meaningful, the scale of measurement is considered a ordinal scale.

Ordinal scale of measurement

- Data exhibits properties of nominal data and the order or rank of data is meaningful, the scale of measurement is considered a ordinal scale.
- Each customer who visits a restaurant provides a service rating of excellent, good, or poor.
  - The data obtained are the labels—excellent, good, or poor—the data have the properties of nominal data.
  - In addition, the data can be ranked, or ordered, with respect to the service quality.

Ordinal scale of measurement

- Data exhibits properties of nominal data and the order or rank of data is meaningful, the scale of measurement is considered a ordinal scale.
- Each customer who visits a restaurant provides a service rating of excellent, good, or poor.
  - The data obtained are the labels—excellent, good, or poor—the data have the properties of nominal data.
  - In addition, the data can be ranked, or ordered, with respect to the service quality.
- Ordinal – name categories that can be ordered

Interval scale of measurement

- If the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure, then the scale of measurement is interval scale.

Interval scale of measurement

- If the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure, then the scale of measurement is interval scale.
- Interval data are always numeric. Can find out difference between any two values.

Interval scale of measurement

- If the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure, then the scale of measurement is interval scale.
- Interval data are always numeric. Can find out difference between any two values.
- Ratios of values have no meaning here because the value of zero is arbitrary.

Interval scale of measurement

- If the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure, then the scale of measurement is interval scale.
- Interval data are always numeric. Can find out difference between any two values.
- Ratios of values have no meaning here because the value of zero is arbitrary.
- Interval:

  numerical values that can be added/subtracted (no absolute zero)

Example: temperature

- Suppose the response to a question on how hot the day is comfortable and uncomfortable, then the temperature as a variable is nominal.

Example: temperature

- Suppose the response to a question on how hot the day is comfortable and uncomfortable, then the temperature as a variable is nominal.
- Suppose the answer to measuring the temperature of a liquid is cold, warm, hot - the variable is ordinal.

Example: temperature

- Suppose the response to a question on how hot the day is comfortable and uncomfortable, then the temperature as a variable is nominal.
- Suppose the answer to measuring the temperature of a liquid is cold, warm, hot - the variable is ordinal.
- Example: Consider a AC room where temperature is set at 20°C and the temperature outside the room is 40°C. It is correct to say that the difference in temperature is 20°C, but it is incorrect to say that the outdoors is twice as hot as indoors.

Example: temperature

- Suppose the response to a question on how hot the day is comfortable and uncomfortable, then the temperature as a variable is nominal.
- Suppose the answer to measuring the temperature of a liquid is cold, warm, hot - the variable is ordinal.
- Example: Consider a AC room where temperature is set at 20°C and the temperature outside the room is 40°C. It is correct to say that the difference in temperature is 20°C, but it is incorrect to say that the outdoors is twice as hot as indoors.
- Temperature in degrees Fahrenheit or degrees centigrade is an interval variable. No absolute zero.

	Celsius	Fahrenheit
Freezing point	0	32
Boiling point	100	212

Ratio scale of measurement

- If the data have all the properties of interval data and the ratio of two values is meaningful, then the scale of measurement is ratio scale.

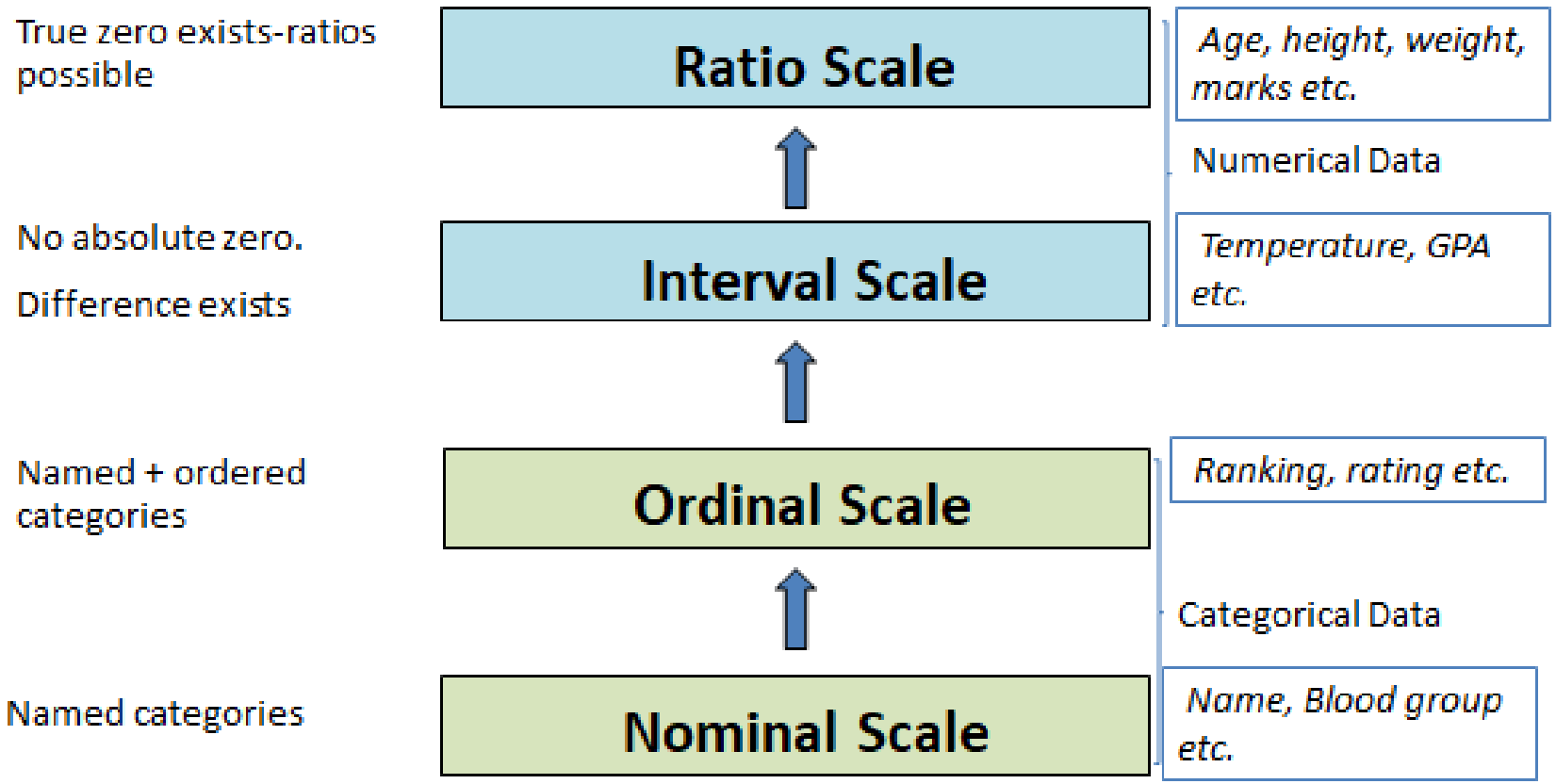
Ratio scale of measurement

- If the data have all the properties of interval data and the ratio of two values is meaningful, then the scale of measurement is ratio scale.
- Example: height, weight, age, marks, etc.

Ratio scale of measurement

- If the data have all the properties of interval data and the ratio of two values is meaningful, then the scale of measurement is ratio scale.
- Example: height, weight, age, marks, etc.
- Ratio: numerical values that can be added, subtracted, multiplied or divided (makes ratio comparisons possible)

Summary


True zero exists-ratios possible
Ratio Scale
Age, height, weight, marks etc.
Numerical Data
No absolute zero.
Difference exists
Interval Scale
Temperature, GPA etc.
Named + ordered categories
Ordinal Scale
Ranking, rating etc.
Categorical Data
Named categories
Nominal Scale
Name, Blood group etc.