# 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.

# What is Statistics?

### Definition
*Statistics[1] is the art of learning from data. It is concerned with the collection of data, their subsequent description, and their analysis, which often leads to the drawing of conclusions.*

---

[1]Ross, Sheldon M. Introductory statistics. Academic Press, 2017.

# Major branches of statistics

1. Description

### Definition
*The part of statistics concerned with the description and summarization of data is called descriptive statistics.*

2. Inference

### Definition
*The part of statistics concerned with the drawing of conclusions from data is called inferential statistics.*

▶ To be able to draw a conclusion from the data, we must take into account the possibility of chance- introduction to probability.

## Population and sample

Suppose we are interested in knowing

▶ The percentage of all students in India who have passed their Class 12 exams and study engineering.

▶ The prices of all houses in Tamil Nadu.

▶ The total sales of all cars in India in the year 2019.

▶ The age distribution of people who visit a city Mall in a particular month.

## Population and sample

### Definition

*The total collection of all the elements that we are interested in is called a population.*

# Population and sample

### Definition
*The total collection of all the elements that we are interested in is called a population.*

### Definition
*A subgroup of the population that will be studied in detail is called a sample.*

# Purpose of statistical analysis

---

[2]Weiss, Neil A. Introductory Statistics: Pearson New International Edition.
Pearson Education Limited, 2014.

## Purpose of statistical analysis

▶ If the purpose of the analysis[2] is to examine and explore information for its own intrinsic interest only, the study is descriptive.

---

[2]Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.

## Purpose of statistical analysis

- ▶ If the purpose of the analysis[2] is to examine and explore information for its own intrinsic interest only, the study is descriptive.

- ▶ If the information is obtained from a sample of a population and the purpose of the study is to use that information to draw conclusions about the population, the study is inferential.

---

[2]Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.

# Purpose of statistical analysis

- ▶ If the purpose of the analysis[2] is to examine and explore information for its own intrinsic interest only, the study is descriptive.

- ▶ If the information is obtained from a sample of a population and the purpose of the study is to use that information to draw conclusions about the population, the study is inferential.

- ▶ A descriptive study may be performed either on a sample or on a population.

---

[2]Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.

## Purpose of statistical analysis

- ▶ If the purpose of the analysis[2] is to examine and explore information for its own intrinsic interest only, the study is descriptive.

- ▶ If the information is obtained from a sample of a population and the purpose of the study is to use that information to draw conclusions about the population, the study is inferential.

- ▶ A descriptive study may be performed either on a sample or on a population.

- ▶ When an inference is made about the population, based on information obtained from the sample, does the study become inferential.

---

[2]Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.

# Summary

# Summary

► Descriptive statistics

# Summary

- ▶ | Descriptive statistics |
- ▶ | Inferential statistics |

# Summary

- ▶ Descriptive statistics
- ▶ Inferential statistics
- ▶ Population and sample