



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

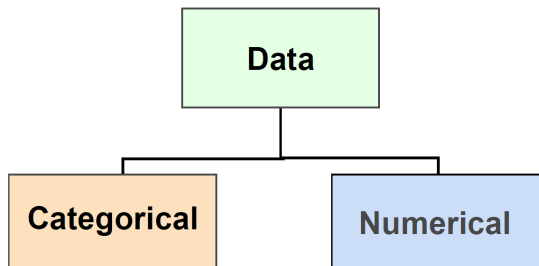
Understanding data

Classification of data

- Categorical and numerical

- Cross-sectional versus time-series data

Categorical and numerical



Categorical and numerical variables

- ▶ Categorical data
 - ▶ Also called qualitative variables.
 - ▶ Identify group membership
- ▶ Numerical data
 - ▶ Also called quantitative variables.
 - ▶ Describe numerical properties of cases
 - ▶ Have measurement units
- ▶ Measurement units: Scale that defines the meaning of numerical data, such as weights measured in kilograms, prices in rupees, heights in centimeters, etc.
 - ▶ The data that make up a numerical variable in a data table must share a common unit.

Cross-sectional and time-series data

- ▶ Time series - data recorded over time
- ▶ Timeplot – graph of a time series showing values in chronological order
- ▶ Cross-sectional - data observed at the same time

- └ Classification of data
 - └ Cross-sectional versus time-series data

Time-series data- Example

<i>Date</i>	<i>Potato</i>		
	<i>Qty(KG)</i>	<i>cost (Rs.)</i>	<i>Selling price(Rs.)</i>
01-Mar	0	21	24
02-Mar	1350	20.05	24
03-Mar	675	20.5	24
04-Mar	0	NA	NA
05-Mar	675	20.8	24
06-Mar	675	21.25	24
08-Mar	20	20.5	24
09-Mar	900	20.5	24
10-Mar	900	20.5	24
11-Mar	0	NA	NA
12-Mar	900	20.3	24
13-Mar	1125	19.4	22
15-Mar	1125	18.8	22
16-Mar	1125	19.4	22
17-Mar	1125	19.25	22
18-Mar	1125	20.3	24
19-Mar	1125	19.8	24
20-Mar	675	21.25	24
22-Mar	675	20.5	24
23-Mar	0	NA	NA
24-Mar	0	NA	NA
25-Mar	675	19.6	24
26-Mar	675	19.7	24
27-Mar	1125	19.3	24
29-Mar	540	20.6	26
30-Mar	0		28

Summary

- ▶ Classify data as categorical or numerical.
- ▶ For numerical data, find out unit of measurement.
- ▶ Check whether data is collected at a point of time (cross-sectional data) or over time (time-series data).