



**IIT Madras**  
ONLINE DEGREE

# Statistics for Data Science -1

## Describing Categorical Data- Single Variable

Usha Mohan

Indian Institute of Technology Madras

# Review

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

## Frequency distributions

### Relative frequency distributions

# Frequency distributions

## Definition

A *frequency distribution*<sup>1</sup> of qualitative data is a listing of the distinct values and their frequencies.

---

<sup>1</sup>Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.

# Frequency distributions

## Definition

A *frequency distribution*<sup>1</sup> of qualitative data is a listing of the distinct values and their frequencies.

Each row of a frequency table lists a category along with the number of cases in this category.

---

<sup>1</sup>Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.

## Example

Construct a frequency table for the given data

1. A,A,B,C,A,D,A,B,D,C
2. A,A,B,C,A,D,A,B,D,C,A,B,C,D,A
3. A,A,B,C,A,A,B,B,D,C,A,B,C,D,B
4. A,A,B,C,A,D,A,B,D,C, A,B,C,D,A,C,D,D

## Construct a frequency distribution

The steps to construct a frequency distribution<sup>2</sup>

- Step 1** List the distinct values of the observations in the data set in the first column of a table.
- Step 2** For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.
- Step 3** Count the tallies for each distinct value and record the totals in the third column of the table.

---

<sup>2</sup>Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.



## Example

1. A,A,B,C,A,D,A,B,D,C

Category	Tally mark	Frequency
A		
B		
C		
D		
<b>Total</b>		

## Example

1. A,A,B,C,A,D,A,B,D,C

Category	Tally mark	Frequency
A		
B		
C		
D		
<b>Total</b>		

2. A,A,B,C,A,D,A,B,D,C, A,B,C,D,A

Category	Tally mark	Frequency
A		
B		
C		
D		
<b>Total</b>		

## Example

3. A,B,B,C,A,D,B,B,D,C, A,B,C,D,B

Category	Tally mark	Frequency
A		
B		
C		
D		
<b>Total</b>		

## Example

3. A,B,B,C,A,D,B,B,D,C, A,B,C,D,B

Category	Tally mark	Frequency
A		
B		
C		
D		
<b>Total</b>		

4. A,A,B,C,A,D,A,B,D,C, A,B,C,D,A,C,D,D

Category	Tally mark	Frequency
A		
B		
C		
D		
<b>Total</b>		

## Frequency table in a googlesheet

- Step 1 Select/Highlight the cells having data you want to visualize.
- Step 2 In the Formatting bar click on the Data option.
- Step 3 In the Data option go to Pivot Table option and create a new sheet.
- Step 4 After creating Pivot Table go in Pivot Table Editor and in that first add rows and then values.

# Relative frequency

## Definition

*The ratio of the frequency to the total number of observations is called **relative frequency***

- The steps to construct a relative frequency distribution
  - Step 1** Obtain a frequency distribution of the data.
  - Step 2** Divide each frequency by the total number of observations.

## Example

1. A,A,B,C,A,D,A,B,D,C

Category	Tally mark	Frequency	Relative frequency
A		4	
B		2	
C		2	
D		2	
<b>Total</b>		10	

## Example

1. A,A,B,C,A,D,A,B,D,C

Category	Tally mark	Frequency	Relative frequency
A		4	
B		2	
C		2	
D		2	
<b>Total</b>		10	

2. A,A,B,C,A,D,A,B,D,C, A,B,C,D,A

Category	Tally mark	Frequency	Relative frequency
A		6	
B		3	
C		3	
D		3	
<b>Total</b>		15	



## Why relative frequency?

- ▶ For comparing two data sets.
- ▶ Because relative frequencies always fall between 0 and 1, they provide a standard for comparison.

## Example

3. A,B,B,C,A,D,B,B,D,C, A,B,C,D,B

Category	Tally mark	Frequency	Relative frequency
A		3	
B		6	
C		3	
D		3	
<b>Total</b>		15	

## Example

3. A,B,B,C,A,D,B,B,D,C, A,B,C,D,B

Category	Tally mark	Frequency	Relative frequency
A		3	
B		6	
C		3	
D		3	
<b>Total</b>		15	

4. A,A,B,C,A,D,A,B,D,C, A,B,C,D,A,C,D,D

Category	Tally mark	Frequency	Relative frequency
A		6	
B		3	
C		4	
D		5	
<b>Total</b>		18	

## Sectional summary

1. Constructing a frequency table.
2. Notion of relative frequency and constructing a relative frequency table.