IIT Madras

ONLINE DEGREE

Statistics for Data Science -1

Lecture 3.2: Describing Numerical Data- Mean

Usha Mohan

Indian Institute of Technology Madras

Descriptive measures

Descriptive measures

- The objective is to develop measures that can be used to summarize a data set.

Descriptive measures

- The objective is to develop measures that can be used to summarize a data set.
- These descriptive measures are quantities whose values are determined by the data.

Descriptive measures

Most commonly used descriptive measures can be categorized as

Descriptive measures

Most commonly used descriptive measures can be categorized as

- Measures of central tendency: These are measures that indicate the most typical value or center of a data set.

Descriptive measures

Most commonly used descriptive measures can be categorized as

- Measures of central tendency: These are measures that indicate the most typical value or center of a data set.
- Measures of dispersion: These measures indicate the variability or spread of a dataset.

The Mean

The most commonly used measure of central tendency is the mean.

Definition

The mean of a data set is the sum of the observations divided by the number of observations.

The Mean

The most commonly used measure of central tendency is the mean.

Definition

The mean of a data set is the sum of the observations divided by the number of observations.

- The mean is usually referred to as average.

The Mean

The most commonly used measure of central tendency is the mean.

Definition

The mean of a data set is the sum of the observations divided by the number of observations.

- The mean is usually referred to as average.
- Arithmetic average; divide the sum of the values by the number of values (another typical value)

The Mean

The most commonly used measure of central tendency is the mean.

Definition

The mean of a data set is the sum of the observations divided by the number of observations.

- The mean is usually referred to as average.
- Arithmetic average; divide the sum of the values by the number of values (another typical value)
- For discrete observations:

The Mean

The most commonly used measure of central tendency is the mean.

Definition

The mean of a data set is the sum of the observations divided by the number of observations.

- The mean is usually referred to as average.
- Arithmetic average; divide the sum of the values by the number of values (another typical value)
- For discrete observations:
  - Sample mean: $\bar{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

The Mean

The most commonly used measure of central tendency is the mean.

Definition

The mean of a data set is the sum of the observations divided by the number of observations.

- The mean is usually referred to as average.
- Arithmetic average; divide the sum of the values by the number of values (another typical value)
- For discrete observations:
  - Sample mean: $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$
  - Population mean: $\mu = \frac{x_1 + x_2 + \ldots + x_N}{N}$

Example

1. 2, 12, 5, 7, 6, 7, 3;

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7}$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} = \frac{135}{7} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} = \frac{135}{7} = 19.285$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} = \frac{135}{7} = 19.285$
3. 2, 105, 5, 7, 6, 3

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} = \frac{135}{7} = 19.285$
3. 2, 105, 5, 7, 6, 3 $\bar{x} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} = \frac{135}{7} = 19.285$
3. 2, 105, 5, 7, 6, 3 $\bar{x} = \frac{2+105+5+7+6+3}{6} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} = \frac{135}{7} = 19.285$
3. 2, 105, 5, 7, 6, 3 $\bar{x} = \frac{2+105+5+7+6+3}{6} = \frac{128}{6} =$

Example

1. 2, 12, 5, 7, 6, 7, 3;
   $\bar{x} = \frac{2+12+5+7+6+7+3}{7} = \frac{42}{7} = 6$
2. 2, 105, 5, 7, 6, 7, 3 $\bar{x} = \frac{2+105+5+7+6+7+3}{7} = \frac{135}{7} = 19.285$
3. 2, 105, 5, 7, 6, 3 $\bar{x} = \frac{2+105+5+7+6+3}{6} = \frac{128}{6} = 21.33$

Example

- The marks obtained by ten students in an exam is
68, 79, 38, 68, 35, 70, 61, 47, 58, 66

Example

- The marks obtained by ten students in an exam is 68, 79, 38, 68, 35, 70, 61, 47, 58, 66
- The sample mean is

$$\frac{68+79+38+68+35+70+61+47+58+66}{10} =$$

Example

- The marks obtained by ten students in an exam is 68, 79, 38, 68, 35, 70, 61, 47, 58, 66
- The sample mean is

$$\frac{68+79+38+68+35+70+61+47+58+66}{10} = \frac{590}{10} =$$

Example

- The marks obtained by ten students in an exam is 68, 79, 38, 68, 35, 70, 61, 47, 58, 66
- The sample mean is

$$\frac{68+79+38+68+35+70+61+47+58+66}{10} = \frac{590}{10} = 59$$

Mean for grouped data: discrete single value data

- The following data is the response from 15 individuals.
  2, 1, 3, 4, 5, 2, 3, 3, 3, 4, 4, 1, 2, 3, 4

Mean for grouped data: discrete single value data

- The following data is the response from 15 individuals.
  2, 1, 3, 4, 5, 2, 3, 3, 3, 4, 4, 1, 2, 3, 4
- $\bar{x} = \dfrac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{n}$

Mean for grouped data: discrete single value data

- The following data is the response from 15 individuals.
  2, 1, 3, 4, 5, 2, 3, 3, 3, 4, 4, 1, 2, 3, 4
- $\bar{x} = \dfrac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{n}$

Value(x_i)	Tally mark	Frequency(f_i)	f_ix_i
1	||	2	2
2	|||	3	6
3	$\cancel{\|\|\|\|}$	5	15
4	||||	4	16
5	|	1	5
Total		15	44

Mean=

Mean for grouped data: discrete single value data

- The following data is the response from 15 individuals.
  2, 1, 3, 4, 5, 2, 3, 3, 3, 4, 4, 1, 2, 3, 4
- $\bar{x} = \dfrac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{n}$

Value(x_i)	Tally mark	Frequency(f_i)	f_ix_i
1	||	2	2
2	|||	3	6
3	~~||||~~	5	15
4	||||	4	16
5	|	1	5
Total		15	44

Mean$=\frac{44}{15} =$

Mean for grouped data: discrete single value data

- The following data is the response from 15 individuals.
  2, 1, 3, 4, 5, 2, 3, 3, 3, 4, 4, 1, 2, 3, 4
- $\bar{x} = \dfrac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{n}$

Value(x_i)	Tally mark	Frequency(f_i)	f_ix_i
1	||	2	2
2	|||	3	6
3	~~||||~~	5	15
4	||||	4	16
5	|	1	5
Total		15	44

Mean$=\frac{44}{15} = 2.93$

Mean for grouped data: continuous data

Mean for grouped data: continuous data

- $\bar{x} = \dfrac{f_1 m_1 + f_2 m_2 + \ldots + f_n m_n}{n}$

Mean for grouped data: continuous data

- $\bar{x} = \dfrac{f_1 m_1 + f_2 m_2 + \ldots + f_n m_n}{n}$

Class interval	Tally mark	Frequency(f_i)	Mid point(m_i)	$f_i m_i$
30-40	$\vert\vert\vert$	3	35	105
40-50	$\cancel{\vert\vert\vert\vert}\ \vert$	6	45	270
50-60	$\cancel{\vert\vert\vert\vert}\ \cancel{\vert\vert\vert\vert}\ \cancel{\vert\vert\vert\vert}\ \vert\vert\vert$	18	55	990
60-70	$\cancel{\vert\vert\vert\vert}\ \cancel{\vert\vert\vert\vert}\ \cancel{\vert\vert\vert\vert}\ \vert\vert$	17	65	1105
70-80	$\vert\vert\vert\vert$	4	75	300
80-90	$\vert\vert$	2	85	170
Total		50		2940

Mean for grouped data: continuous data

- $\bar{x} = \dfrac{f_1 m_1 + f_2 m_2 + \ldots + f_n m_n}{n}$

Class interval	Tally mark	Frequency(f_i)	Mid point(m_i)	$f_i m_i$
30-40	|||	3	35	105
40-50	~~||||~~ |	6	45	270
50-60	~~||||~~ ~~||||~~ ~~||||~~ |||	18	55	990
60-70	~~||||~~ ~~||||~~ ~~||||~~ ||	17	65	1105
70-80	||||	4	75	300
80-90	||	2	85	170
Total		50		2940

- Average $= \frac{2940}{50} = 58.8$.

Mean for grouped data: continuous data

- $\bar{x} = \dfrac{f_1 m_1 + f_2 m_2 + \ldots + f_n m_n}{n}$

Class interval	Tally mark	Frequency(f_i)	Mid point(m_i)	$f_i m_i$
30-40	|||	3	35	105
40-50	~~||||~~ |	6	45	270
50-60	~~||||~~ ~~||||~~ ~~||||~~ |||	18	55	990
60-70	~~||||~~ ~~||||~~ ~~||||~~ ||	17	65	1105
70-80	||||	4	75	300
80-90	||	2	85	170
Total		50		2940

- Average $= \frac{2940}{50} = 58.8$.
- 58.8 is an <u>approximate</u> and not <u>exact</u> value of the mean

Adding a constant

Adding a constant

- Let $y_i = x_i + c$ where c is a constant then $\bar{y} = \bar{x} + c$

Adding a constant

- Let $y_i = x_i + c$ where c is a constant then $\bar{y} = \bar{x} + c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.

Adding a constant

- Let $y_i = x_i + c$ where c is a constant then $\bar{y} = \bar{x} + c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.
  - Suppose the teacher has decided to add 5 marks to each student.
  - Then the data becomes
    73, 84, 43, 73, 40, 75, 66, 52, 63, 71.

Adding a constant

- Let $y_i = x_i + c$ where c is a constant then $\bar{y} = \bar{x} + c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.
  - Suppose the teacher has decided to add 5 marks to each student.
  - Then the data becomes
    73, 84, 43, 73, 40, 75, 66, 52, 63, 71.
  - The mean of the new data set is $\frac{640}{10} = 64 = 59 + 5$

Multiplying a constant

Multiplying a constant

- Let $y_i = x_i c$ where c is a constant then $\bar{y} = \bar{x}c$

Multiplying a constant

- Let $y_i = x_i c$ where c is a constant then $\bar{y} = \bar{x}c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.

Multiplying a constant

- Let $y_i = x_i c$ where c is a constant then $\bar{y} = \bar{x}c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.
  - Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.

Multiplying a constant

- Let $y_i = x_i c$ where c is a constant then $\bar{y} = \bar{x}c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.
  - Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.
  - Then the data becomes
    27.2, 31.6, 15.2, 27.2, 14, 28, 24.4, 18.8, 23.2, 26.4

Multiplying a constant

- Let $y_i = x_i c$ where c is a constant then $\bar{y} = \bar{x}c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.
  - Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.
  - Then the data becomes
    27.2, 31.6, 15.2, 27.2, 14, 28, 24.4, 18.8, 23.2, 26.4
  - The mean of the new data set is $\frac{236}{10} = 23.6 =$

Multiplying a constant

- Let $y_i = x_i c$ where c is a constant then $\bar{y} = \bar{x}c$
- Example: Recall the marks of students
  68, 79, 38, 68, 35, 70, 61, 47, 58, 66.
  - Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.
  - Then the data becomes
    27.2, 31.6, 15.2, 27.2, 14, 28, 24.4, 18.8, 23.2, 26.4
  - The mean of the new data set is $\frac{236}{10} = 23.6 = 59 \times 0.4$

Section summary

1. Mean or average is a measure of central tendency.
2. Compute sample mean for
   2.1 ungrouped data.
   2.2 grouped discrete data.
   2.3 grouped continuous data.
3. Manipulating data
   3.1 Adding a constant to each data point.
   3.2 Multiplying each data point with a constant.