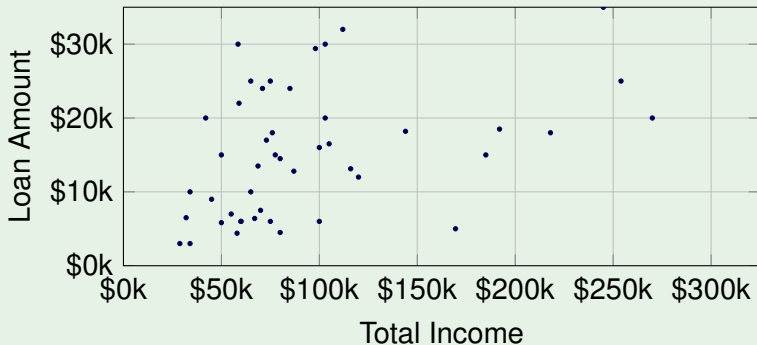


Examining Numerical Data

Colby Community College

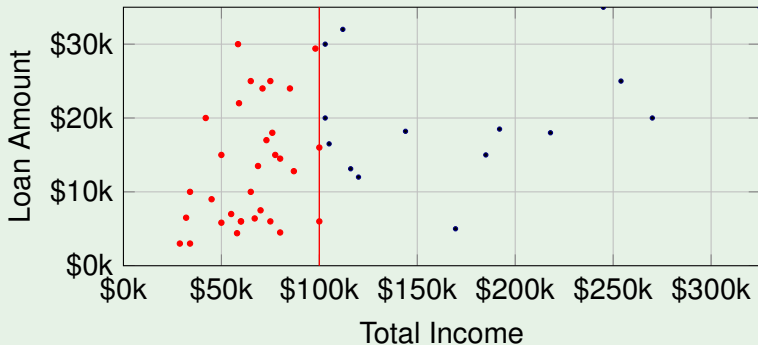
Example 1

Let us consider a scatterplot of borrowers total income and the loan amount from the `loan50` data set.



Example 1

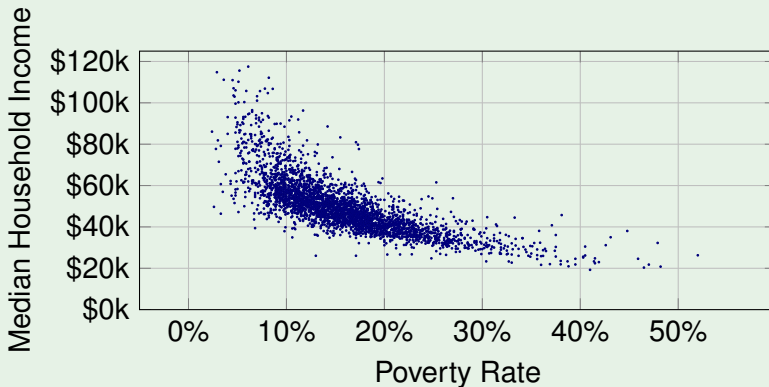
Let us consider a scatterplot of borrowers total income and the loan amount from the `loan50` data set.



We can see that the many of borrowers earn \$100,000 a year or less.

Example 2

Let us consider a scatterplot of borrowers total income and the loan amount from the `loan50` data set.



Example 2

Let us consider a scatterplot of borrowers total income and the loan amount from the `loan50` data set.



It is clear there is a **nonlinear** association between the median household income and the poverty rate.

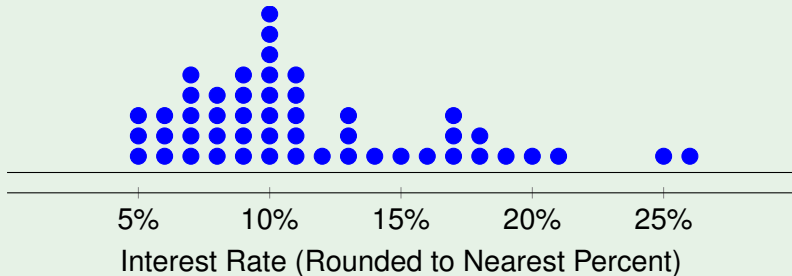
Definition

A **dot plot** is a one-variable scatterplot. Each data value is plotted as a point above a horizontal scale of values. Dots representing equal values are stacked.

Definition

A **dot plot** is a one-variable scatterplot. Each data value is plotted as a point above a horizontal scale of values. Dots representing equal values are stacked.

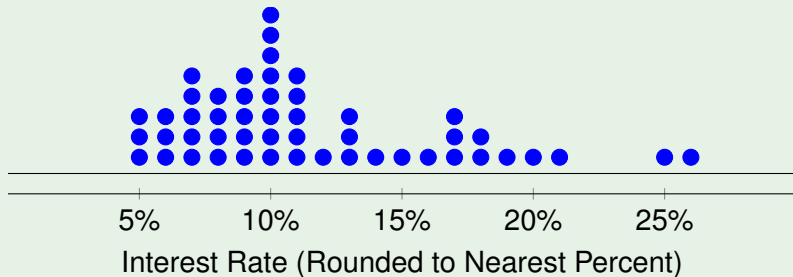
Example 3



Definition

A **dot plot** is a one-variable scatterplot. Each data value is plotted as a point above a horizontal scale of values. Dots representing equal values are stacked.

Example 3



Note

Dot plots work best with integer data. It is common to round decimals before building a dot plot.

Definition

A **parameter** is a numerical measurement describing some characteristic of a population.

Definition

A **parameter** is a numerical measurement describing some characteristic of a population.

Definition

A **statistic** is a numerical measurement describing some characteristic of a sample.

Definition

A **parameter** is a numerical measurement describing some characteristic of a population.

Definition

A **statistic** is a numerical measurement describing some characteristic of a sample.

Note

Parameter and population both start with a “P.”
Statistic and sample both start with a “S.”

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

Note

The mean is also known as the **average**.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

Note

The mean is also known as the **average**.

Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

Note

The mean is also known as the **average**.

Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.
- The mean of a data set uses every data value.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

Note

The mean is also known as the **average**.

Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.
- The mean of a data set uses every data value.
- A disadvantage of the mean is that just one extreme value can change the value of the mean substantially.

Common Notation

Sample statistics are usually represented by English letters, such as \bar{x} , while population parameters are usually represented by Greek letters, such as μ .

Common Notation

Sample statistics are usually represented by English letters, such as \bar{x} , while population parameters are usually represented by Greek letters, such as μ .

Σ denotes the sum of a set of data values.

Common Notation

Sample statistics are usually represented by English letters, such as \bar{x} , while population parameters are usually represented by Greek letters, such as μ .

Σ denotes the sum of a set of data values.

x is used as a placeholder for the variable of interest.

Common Notation

Sample statistics are usually represented by English letters, such as \bar{x} , while population parameters are usually represented by Greek letters, such as μ .

Σ denotes the sum of a set of data values.

x is used as a placeholder for the variable of interest.

n represents the number of data values in a sample.

Common Notation

Sample statistics are usually represented by English letters, such as \bar{x} , while population parameters are usually represented by Greek letters, such as μ .

Σ denotes the sum of a set of data values.

x is used as a placeholder for the variable of interest.

n represents the number of data values in a sample.

N represents the number of data values in a population.

Common Notation

Sample statistics are usually represented by English letters, such as \bar{x} , while population parameters are usually represented by Greek letters, such as μ .

\sum denotes the sum of a set of data values.

x is used as a placeholder for the variable of interest.

n represents the number of data values in a sample.

N represents the number of data values in a population.

$\bar{x} = \frac{\sum x}{n}$ is the mean of a set of sample values.

Common Notation

Sample statistics are usually represented by English letters, such as \bar{x} , while population parameters are usually represented by Greek letters, such as μ .

Σ denotes the sum of a set of data values.

x is used as a placeholder for the variable of interest.

n represents the number of data values in a sample.

N represents the number of data values in a population.

$\bar{x} = \frac{\Sigma x}{n}$ is the mean of a set of sample values.

$\mu = \frac{\Sigma x}{N}$ is the mean of all values in a population.

Example 4

Suppose we measure the of data speeds of smartphones from the four major carriers. The table contains five data speeds, in megabits per second (Mbps), from this data set.

Carrier	Verizon	Verizon	Verizon	Verizon	Verizon
Mbps	38.5	55.6	22.4	14.1	23.1

Example 4

Suppose we measure the of data speeds of smartphones from the four major carriers. The table contains five data speeds, in megabits per second (Mbps), from this data set.

Carrier	Verizon	Verizon	Verizon	Verizon	Verizon
Mbps	38.5	55.6	22.4	14.1	23.1

The mean is

$$\bar{x} = \frac{\sum x}{n}$$

Example 4

Suppose we measure the of data speeds of smartphones from the four major carriers. The table contains five data speeds, in megabits per second (Mbps), from this data set.

Carrier	Verizon	Verizon	Verizon	Verizon	Verizon
Mbps	38.5	55.6	22.4	14.1	23.1

The mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5}$$

Example 4

Suppose we measure the of data speeds of smartphones from the four major carriers. The table contains five data speeds, in megabits per second (Mbps), from this data set.

Carrier	Verizon	Verizon	Verizon	Verizon	Verizon
Mbps	38.5	55.6	22.4	14.1	23.1

The mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = \frac{153.7}{5}$$

Example 4

Suppose we measure the of data speeds of smartphones from the four major carriers. The table contains five data speeds, in megabits per second (Mbps), from this data set.

Carrier	Verizon	Verizon	Verizon	Verizon	Verizon
Mbps	38.5	55.6	22.4	14.1	23.1

The mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = \frac{153.7}{5} = 30.74 \text{ Mbps}$$

Example 4

Suppose we measure the of data speeds of smartphones from the four major carriers. The table contains five data speeds, in megabits per second (Mbps), from this data set.

Carrier	Verizon	Verizon	Verizon	Verizon	Verizon
Mbps	38.5	55.6	22.4	14.1	23.1

The mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = \frac{153.7}{5} = 30.74 \text{ Mbps}$$

Note

Round statistics and parameters to one more decimal place than found in the data.

Note

It is common to mark the mean on a dot plot.

Note

It is common to mark the mean on a dot plot.

Example 5

The mean of `interest_rate` is: (Do not round the data values.)

$$\bar{x} = \frac{\left(\begin{array}{l} 5.31\% + 5.31\% + 5.32\% + 6.08\% + 6.08\% + 6.08\% + 6.71\% + 6.71\% + 7.34\% \\ +7.35\% + 7.35\% + 7.96\% + 7.96\% + 7.96\% + 7.97\% + 9.43\% + 9.43\% + 9.44\% \\ +9.44\% + 9.44\% + 9.92\% + 9.92\% + 9.92\% + 9.92\% + 9.93\% + 9.93\% + 10.42\% \\ +10.42\% + 10.9\% + 10.9\% + 10.91\% + 10.91\% + 10.91\% + 11.98\% + 12.62\% \\ +12.62\% + 12.62\% + 14.08\% + 15.04\% + 16.02\% + 17.09\% + 17.09\% + 17.09\% \\ +18.06\% + 18.45\% + 19.42\% + 20\% + 21.45\% + 24.85\% + 26.3\% \end{array} \right)}{50} = 11.567\%$$

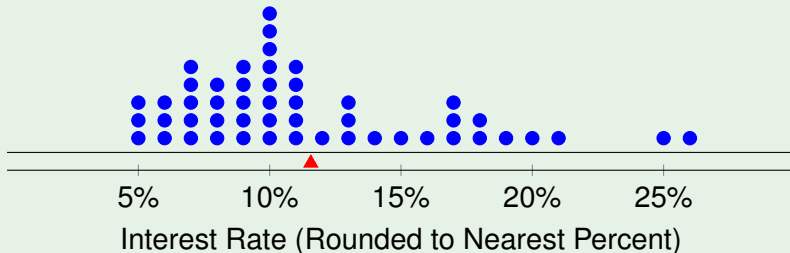
Note

It is common to mark the mean on a dot plot.

Example 5

The mean of `interest_rate` is: (Do not round the data values.)

$$\bar{x} = \frac{\left(\begin{array}{l} 5.31\% + 5.31\% + 5.32\% + 6.08\% + 6.08\% + 6.08\% + 6.71\% + 6.71\% + 7.34\% \\ +7.35\% + 7.35\% + 7.96\% + 7.96\% + 7.96\% + 7.97\% + 9.43\% + 9.43\% + 9.44\% \\ +9.44\% + 9.44\% + 9.92\% + 9.92\% + 9.92\% + 9.92\% + 9.93\% + 9.93\% + 10.42\% \\ +10.42\% + 10.9\% + 10.9\% + 10.91\% + 10.91\% + 10.91\% + 11.98\% + 12.62\% \\ +12.62\% + 12.62\% + 14.08\% + 15.04\% + 16.02\% + 17.09\% + 17.09\% + 17.09\% \\ +18.06\% + 18.45\% + 19.42\% + 20\% + 21.45\% + 24.85\% + 26.3\% \end{array} \right)}{50} = 11.567\%$$



Definition

The **median** of a data set is the measure of center that is the middle value when the original data values are arranged in order of increasing magnitude.

Definition

The **median** of a data set is the measure of center that is the middle value when the original data values are arranged in order of increasing magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)

Definition

The **median** of a data set is the measure of center that is the middle value when the original data values are arranged in order of increasing magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)
- The median does not directly use every data value.

Definition

The **median** of a data set is the measure of center that is the middle value when the original data values are arranged in order of increasing magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)
- The median does not directly use every data value.

Notation

The median of a sample is denoted \tilde{x} .

Definition

The **median** of a data set is the measure of center that is the middle value when the original data values are arranged in order of increasing magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)
- The median does not directly use every data value.

Notation

The median of a sample is denoted \tilde{x} .

Procedure

- 1 Sort the values.
- 2
 - If the number of data values is odd, the median is the number located in the exact middle of the sorted list.
 - If the number of data values is even, the median is found by

Example 6

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1
------	------	------	------	------

Example 6

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1
------	------	------	------	------

First sort the data values.

14.1	22.4	23.1	38.5	55.6
------	------	------	------	------

Example 6

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1
------	------	------	------	------

First sort the data values.

14.1	22.4	23.1	38.5	55.6
------	------	------	------	------

We have 5 data values so the median is 23.1 Mbps.

Example 6

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1
------	------	------	------	------

First sort the data values.

14.1	22.4	23.1	38.5	55.6
------	------	------	------	------

We have 5 data values so the median is 23.1 Mbps.

Note

This different than the mean 30.74 Mbps.

Example 7

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains six data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1	24.5
------	------	------	------	------	------

Example 7

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains six data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1	24.5
------	------	------	------	------	------

First sort the data values.

14.1	22.4	23.1	24.5	38.5	55.6
------	------	------	------	------	------

Example 7

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains six data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1	24.5
------	------	------	------	------	------

First sort the data values.

14.1	22.4	23.1	24.5	38.5	55.6
------	------	------	------	------	------

We have 6 data values so the median is $\frac{23.1 + 24.5}{2} = 23.80$ Mbps.