Measures of Relative Standing and Boxplots

Colby Community College

Definition

A z-**score** is the number of standard deviations that a given value is above or below the mean.

Definition

A z-**score** is the number of standard deviations that a given value is above or below the mean.

Formula

The z-score for a given value x is calculated as follows.

Sample: $z = \dfrac{x - \bar{x}}{s}$

Population: $z = \dfrac{x - \mu}{\sigma}$

Definition

A z-**score** is the number of standard deviations that a given value is above or below the mean.

Formula

The z-score for a given value x is calculated as follows.

Sample: $z = \dfrac{x - \bar{x}}{s}$

Population: $z = \dfrac{x - \mu}{\sigma}$

Properties

- z-scores are expressed as numbers with no units.

Definition

A z-**score** is the number of standard deviations that a given value is above or below the mean.

Formula

The z-score for a given value x is calculated as follows.

Sample: $z = \dfrac{x - \bar{x}}{s}$

Population: $z = \dfrac{x - \mu}{\sigma}$

Properties

- z-scores are expressed as numbers with no units.
- A data value is *significantly low* if $z \leq -2$.

Definition

A z-**score** is the number of standard deviations that a given value is above or below the mean.

Formula

The z-score for a given value x is calculated as follows.

Sample: $z = \dfrac{x - \bar{x}}{s}$

Population: $z = \dfrac{x - \mu}{\sigma}$

Properties

- z-scores are expressed as numbers with no units.
- A data value is *significantly low* if $z \leq -2$.
- A data value is *significantly high* if $z \geq 2$.

Definition

A z-**score** is the number of standard deviations that a given value is above or below the mean.

Formula

The z-score for a given value x is calculated as follows.

Sample: $z = \dfrac{x - \bar{x}}{s}$

Population: $z = \dfrac{x - \mu}{\sigma}$

Properties

- z-scores are expressed as numbers with no units.
- A data value is *significantly low* if $z \leq -2$.
- A data value is *significantly high* if $z \geq 2$.
- If a data value is less than the mean, its z-score will be negative.

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s}$$

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3152.0}{693.4}$$

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3152.0}{693.4} = 1.22$$

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3152.0}{693.4} = 1.22$$

Example 2

The weights of a sample of 106 adult temperature has mean $\bar{x} = 98.20^\circ$F and standard deviation $s = 0.62^\circ$F. What is the z-score of an adult with temperature 96.5°F?

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3152.0}{693.4} = 1.22$$

Example 2

The weights of a sample of 106 adult temperature has mean $\bar{x} = 98.20^\circ$F and standard deviation $s = 0.62^\circ$F. What is the z-score of an adult with temperature 96.5°F?

$$z = \frac{x - \bar{x}}{s}$$

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3152.0}{693.4} = 1.22$$

Example 2

The weights of a sample of 106 adult temperature has mean $\bar{x} = 98.20°$F and standard deviation $s = 0.62°$F. What is the z-score of an adult with temperature $96.5°$F?

$$z = \frac{x - \bar{x}}{s} = \frac{96.5 - 98.20}{0.62}$$

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3152.0}{693.4} = 1.22$$

Example 2

The weights of a sample of 106 adult temperature has mean $\bar{x} = 98.20°$F and standard deviation $s = 0.62°$F. What is the z-score of an adult with temperature $96.5°$F?

$$z = \frac{x - \bar{x}}{s} = \frac{96.5 - 98.20}{0.62} = -2.74$$

Example 1

The weights of a sample of 400 newborn baby weights has mean $\bar{x} = 3152.0$ g and standard deviation $s = 693.4$ g. What is the z-score of a 4000 g baby?

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3152.0}{693.4} = 1.22$$

Example 2

The weights of a sample of 106 adult temperature has mean $\bar{x} = 98.20^\circ$F and standard deviation $s = 0.62^\circ$F. What is the z-score of an adult with temperature 96.5°F?

$$z = \frac{x - \bar{x}}{s} = \frac{96.5 - 98.20}{0.62} = -2.74$$

Rounding

Round z-scores to two decimal places.

Definition

Percentiles are measures of location, denoted $P_1, P_2, \ldots, P_{99}$, which divide a set of data into 100 groups with about 1% of the values in each group.

Definition

Percentiles are measures of location, denoted $P_1, P_2, \ldots, P_{99}$, which divide a set of data into 100 groups with about 1% of the values in each group.

Formula

The process of finding the percentile that corresponds to a particular data value x is given by the following:

$$\text{Percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \cdot 100$$

Definition

Percentiles are measures of location, denoted $P_1, P_2, \ldots, P_{99}$, which divide a set of data into 100 groups with about 1% of the values in each group.

Formula

The process of finding the percentile that corresponds to a particular data value x is given by the following:

$$\text{Percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \cdot 100$$

Rounding

Round percentiles to the nearest whole number.

Example 3

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

38.5	55.6	22.4	14.1	23.1	24.5	6.5	21.5	25.7	14.7
77.8	71.3	43.0	20.2	15.5	13.7	11.1	13.5	10.2	21.1
15.1	14.2	4.5	7.9	9.9	10.3	6.2	17.5	22.2	13.1
18.2	28.5	15.8	15.0	11.1	11.8	16.0	10.9	1.8	34.6
4.6	12.0	11.6	3.6	1.9	7.7	0.8	4.5	1.4	3.2

What percentile is the data value 11.8 Mbps in?

Example 3

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

38.5	55.6	22.4	14.1	23.1	24.5	6.5	21.5	25.7	14.7
77.8	71.3	43.0	20.2	15.5	13.7	11.1	13.5	10.2	21.1
15.1	14.2	4.5	7.9	9.9	10.3	6.2	17.5	22.2	13.1
18.2	28.5	15.8	15.0	11.1	11.8	16.0	10.9	1.8	34.6
4.6	12.0	11.6	3.6	1.9	7.7	0.8	4.5	1.4	3.2

What percentile is the data value 11.8 Mbps in?
There are 20 data values less than 11.8 Mbps.

Example 3

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

38.5	55.6	22.4	14.1	23.1	24.5	6.5	21.5	25.7	14.7
77.8	71.3	43.0	20.2	15.5	13.7	11.1	13.5	10.2	21.1
15.1	14.2	4.5	7.9	9.9	10.3	6.2	17.5	22.2	13.1
18.2	28.5	15.8	15.0	11.1	11.8	16.0	10.9	1.8	34.6
4.6	12.0	11.6	3.6	1.9	7.7	0.8	4.5	1.4	3.2

What percentile is the data value 11.8 Mbps in?

There are 20 data values less than 11.8 Mbps.

$$\text{Percentile of } 11.8 = \frac{20}{50} \cdot 100$$

Example 3

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

38.5	55.6	22.4	14.1	23.1	24.5	6.5	21.5	25.7	14.7
77.8	71.3	43.0	20.2	15.5	13.7	11.1	13.5	10.2	21.1
15.1	14.2	4.5	7.9	9.9	10.3	6.2	17.5	22.2	13.1
18.2	28.5	15.8	15.0	11.1	11.8	16.0	10.9	1.8	34.6
4.6	12.0	11.6	3.6	1.9	7.7	0.8	4.5	1.4	3.2

What percentile is the data value 11.8 Mbps in?

There are 20 data values less than 11.8 Mbps.

$$\text{Percentile of } 11.8 = \frac{20}{50} \cdot 100 = 40$$

A data speed of 11.8 Mbps is in the 40th percentile.

Example 3

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

38.5	55.6	22.4	14.1	23.1	24.5	6.5	21.5	25.7	14.7
77.8	71.3	43.0	20.2	15.5	13.7	11.1	13.5	10.2	21.1
15.1	14.2	4.5	7.9	9.9	10.3	6.2	17.5	22.2	13.1
18.2	28.5	15.8	15.0	11.1	11.8	16.0	10.9	1.8	34.6
4.6	12.0	11.6	3.6	1.9	7.7	0.8	4.5	1.4	3.2

What percentile is the data value 11.8 Mbps in?
There are 20 data values less than 11.8 Mbps.

$$\text{Percentile of } 11.8 = \frac{20}{50} \cdot 100 = 40$$

A data speed of 11.8 Mbps is in the 40th percentile.

Note

This can be interpreted loosely as 40% of Verizon data speeds are slower than 11.8 Mbps and 60% of Verizon data speeds are faster than 11.8 Mbps.

Converting a Percentile to a Data Value

Notation:

- n is the total number of values in the data set.
- k is the percentile being used.
- L is the locator that gives the position of a value.
- P_k is the kth percentile.

Converting a Percentile to a Data Value

Notation:

- n is the total number of values in the data set.
- k is the percentile being used.
- L is the locator that gives the position of a value.
- P_k is the kth percentile.

To find which data value is in the P_k percentile:

1. Sort the data from lowest to highest.

Converting a Percentile to a Data Value

Notation:

- n is the total number of values in the data set.
- k is the percentile being used.
- L is the locator that gives the position of a value.
- P_k is the kth percentile.

To find which data value is in the P_k percentile:

1. Sort the data from lowest to highest.
2. Compute $L = \left(\dfrac{k}{100}\right) n$

Converting a Percentile to a Data Value

Notation:

- n is the total number of values in the data set.
- k is the percentile being used.
- L is the locator that gives the position of a value.
- P_k is the kth percentile.

To find which data value is in the P_k percentile:

1. Sort the data from lowest to highest.
2. Compute $L = \left(\frac{k}{100}\right) n$
3.
   - If L is a whole number, the value of the kth percentile is midway between the Lth value and the next value in the sorted data. Add the Lth value and $(L+1)$th value, then divide by 2.

Converting a Percentile to a Data Value

Notation:

- n is the total number of values in the data set.
- k is the percentile being used.
- L is the locator that gives the position of a value.
- P_k is the kth percentile.

To find which data value is in the P_k percentile:

1. Sort the data from lowest to highest.
2. Compute $L = \left(\dfrac{k}{100}\right) n$
3.
    - If L is a whole number, the value of the kth percentile is midway between the Lth value and the next value in the sorted data. Add the Lth value and $(L+1)$th value, then divide by 2.
    - If L is not a whole number, round L up to the nearest whole number. $P_k$is the Lth data value.

Example 4

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

38.5	55.6	22.4	14.1	23.1	24.5	6.5	21.5	25.7	14.7
77.8	71.3	43.0	20.2	15.5	13.7	11.1	13.5	10.2	21.1
15.1	14.2	4.5	7.9	9.9	10.3	6.2	17.5	22.2	13.1
18.2	28.5	15.8	15.0	11.1	11.8	16.0	10.9	1.8	34.6
4.6	12.0	11.6	3.6	1.9	7.7	0.8	4.5	1.4	3.2

What is the value in the 25th percentile, P_{25}?

Example 4

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

What is the value in the 25th percentile, P_{25}?
First, sort the data.

Example 4

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

What is the value in the 25th percentile, P_{25}?
First, sort the data.
We next need to compute

$$L = \frac{k}{100} \cdot n$$

Example 4

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

What is the value in the 25th percentile, P_{25}?

First, sort the data.

We next need to compute

$$L = \frac{k}{100} \cdot n = \frac{25}{100} \cdot 50$$

Example 4

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

What is the value in the 25th percentile, P_{25}?
First, sort the data.
We next need to compute

$$L = \frac{k}{100} \cdot n = \frac{25}{100} \cdot 50 = 12.5$$

Example 4

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

What is the value in the 25th percentile, P_{25}?

First, sort the data.

We next need to compute

$$L = \frac{k}{100} \cdot n = \frac{25}{100} \cdot 50 = 12.5$$

Since $L = 12.5$ is not a whole number, we round up to 13.

Example 4

The table lists the 50 Verizon airport data speeds, in Mbps, from Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

What is the value in the 25th percentile, P_{25}?
First, sort the data.
We next need to compute

$$L = \frac{k}{100} \cdot n = \frac{25}{100} \cdot 50 = 12.5$$

Since $L = 12.5$ is not a whole number, we round up to 13.
So, P_{25} is the 13th data value, 7.9 Mbps.

Definition

Quartiles are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

Definition

Quartiles are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

Quartile Descriptions

First Quartile, Q_1: Same value as P_{25}. It separates the bottom 25% of the sorted values from the top 75%.

Definition

Quartiles are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

Quartile Descriptions

First Quartile, Q_1: Same value as P_{25}. It separates the bottom 25% of the sorted values from the top 75%.

Second Quartile, Q_2: Same as the P_{50} and the median. It separates the bottom 50% of the sorted values from the top 50%

Definition

Quartiles are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

Quartile Descriptions

First Quartile, Q_1: Same value as P_{25}. It separates the bottom 25% of the sorted values from the top 75%.

Second Quartile, Q_2: Same as the P_{50} and the median. It separates the bottom 50% of the sorted values from the top 50%

Third Quartile, Q_3: Same as P_{75}. It separates the bottom 75% of the sorted values from the top 25%.

Definition

Quartiles are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

Quartile Descriptions

First Quartile, Q_1: Same value as P_{25}. It separates the bottom 25% of the sorted values from the top 75%.

Second Quartile, Q_2: Same as the P_{50} and the median. It separates the bottom 50% of the sorted values from the top 50%

Third Quartile, Q_3: Same as P_{75}. It separates the bottom 75% of the sorted values from the top 25%.

Note

Use the same procedure for calculating percentiles to calculate quartiles.

Definition

The **interquartile range (IQR)** is $Q_3 - Q_1$.

Definition

The **interquartile range** (**IQR**) is $Q_3 - Q_1$.

Outliers

A data value is often considered an outlier if

- the data value is greater than $Q_3 + 1.5 \cdot \mathbf{IQR}$.
- the data value is less than $Q_1 - 1.5 \cdot \mathbf{IQR}$.

Definition

The **interquartile range (IQR)** is $Q_3 - Q_1$.

Outliers

A data value is often considered an outlier if

- the data value is greater than $Q_3 + 1.5 \cdot \mathbf{IQR}$.
- the data value is less than $Q_1 - 1.5 \cdot \mathbf{IQR}$.

Definition

The **semi-interquartile range** is $\frac{Q_3 - Q_1}{2}$.

Definition

The **interquartile range (IQR)** is $Q_3 - Q_1$.

Outliers

A data value is often considered an outlier if

- the data value is greater than $Q_3 + 1.5 \cdot \textbf{IQR}$.
- the data value is less than $Q_1 - 1.5 \cdot \textbf{IQR}$.

Definition

The **semi-interquartile range** is $\frac{Q_3 - Q_1}{2}$.

Definition

The **midquartile** is $\frac{Q_3 + Q_1}{2}$.

Definition

The **interquartile range (IQR)** is $Q_3 - Q_1$.

Outliers

A data value is often considered an outlier if

- the data value is greater than $Q_3 + 1.5 \cdot \mathbf{IQR}$.
- the data value is less than $Q_1 - 1.5 \cdot \mathbf{IQR}$.

Definition

The **semi-interquartile range** is $\frac{Q_3 - Q_1}{2}$.

Definition

The **midquartile** is $\frac{Q_3 + Q_1}{2}$.

Definition

The **10-90 percentile range** is $P_{90} - P_{10}$.

Definition

For a set of data, the **5-number summary** consists of the five values:

1. Minimum
2. Q_1
3. Median (Q_2)
4. Q_3
5. Maximum

Example 5

The table lists, in order from lowest to highest, the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

We can compute the following:

$$\text{minimum} = 0.8$$

Example 5

The table lists, in order from lowest to highest, the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

We can compute the following:

$$\text{minimum} = 0.8$$

$$Q_1 = 7.9$$

Example 5

The table lists, in order from lowest to highest, the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

We can compute the following:

$$\text{minimum} = 0.8$$
$$Q_1 = 7.9$$
$$Q_2 = 13.9$$

Example 5

The table lists, in order from lowest to highest, the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

We can compute the following:

$$\text{minimum} = 0.8$$
$$Q_1 = 7.9$$
$$Q_2 = 13.9$$
$$Q_3 = 21.5$$

Example 5

The table lists, in order from lowest to highest, the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

We can compute the following:

$$\begin{aligned} \text{minimum} &= 0.8 \\ Q_1 &= 7.9 \\ Q_2 &= 13.9 \\ Q_3 &= 21.5 \\ \text{maximum} &= 77.8 \end{aligned}$$

Example 5

The table lists, in order from lowest to highest, the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B.

0.8	1.4	1.8	1.9	3.2	3.6	4.5	4.5	4.6	6.2
6.5	7.7	7.9	9.9	10.2	10.3	10.9	11.1	11.1	11.6
11.8	12.0	13.1	13.5	13.7	14.1	14.2	14.7	15.0	15.1
15.5	15.8	16.0	17.5	18.2	20.2	21.1	21.5	22.2	22.4
23.1	24.5	25.7	28.5	34.6	38.5	43.0	55.6	71.3	77.8

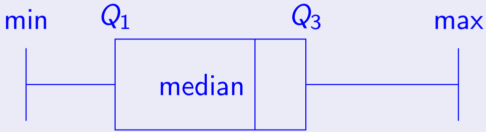
We can compute the following:

$$\text{minimum} = 0.8$$
$$Q_1 = 7.9$$
$$Q_2 = 13.9$$
$$Q_3 = 21.5$$
$$\text{maximum} = 77.8$$

And so, the 5-number summary is 0.8 7.9 13.9 21.5 77.8

Definition

A **boxplot** is a graph of a data set that consists of a line extending from the minimum value to the maximum value, and a box with lines drawn at the first quartile Q_1, the median, and the third quartile Q_3.

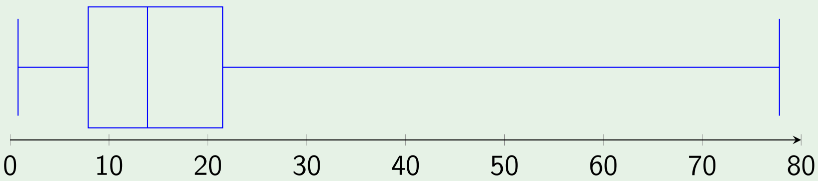


Example 6

The 5-number summary of the the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B is:

0.8 7.9 13.9 21.5 77.8

The boxplot is:

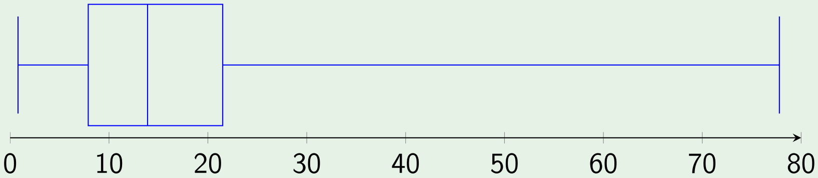


Example 6

The 5-number summary of the the 50 Verizon airport data speeds, in Mbps, from
Data Set 32 in Appendix B is:

$$0.8 \quad 7.9 \quad 13.9 \quad 21.5 \quad 77.8$$

The boxplot is:

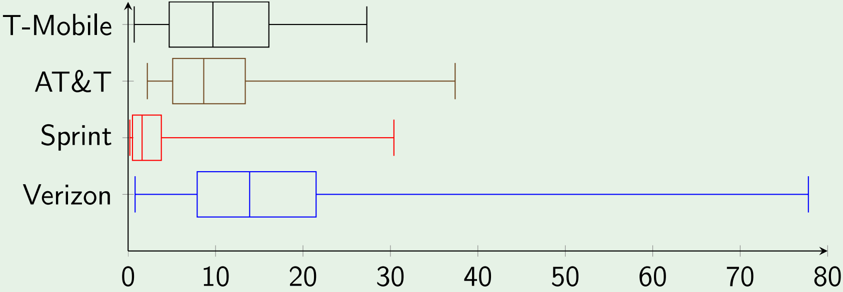


Note

A boxplot can often be used to identify skewness.

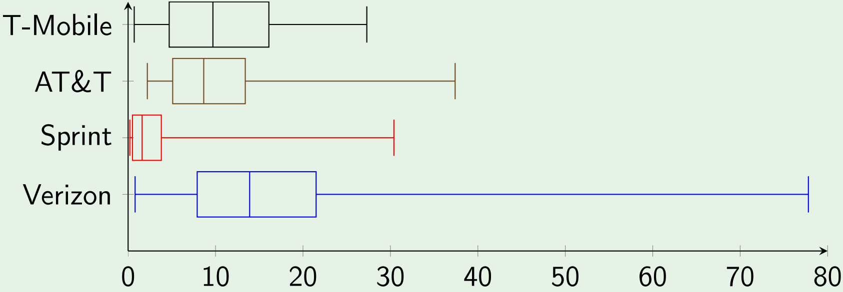
Example 7

We can use boxplots to easily compare the four carriers from Data Set 32 in Appendix B.



Example 7

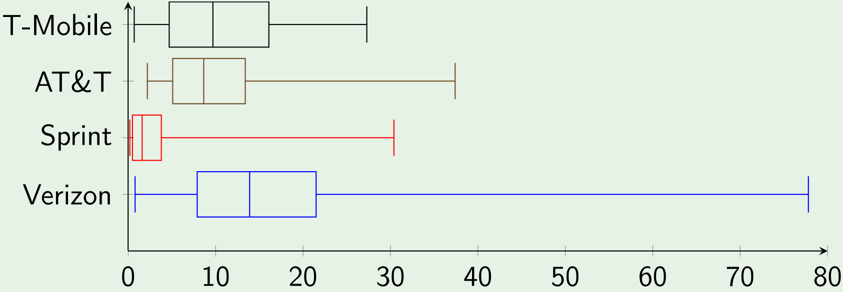
We can use boxplots to easily compare the four carriers from Data Set 32 in Appendix B.



Verizon and T-Mobile are not that different for over 75% of customers.

Example 7

We can use boxplots to easily compare the four carriers from Data Set 32 in Appendix B.



Verizon and T-Mobile are not that different for over 75% of customers.

Outliers

It is important to identify outliers because they can strongly affect the values of important statistics.