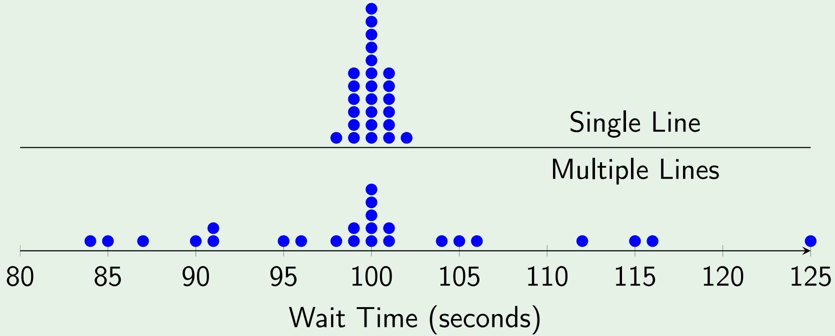
Measures of Variation

Colby Community College

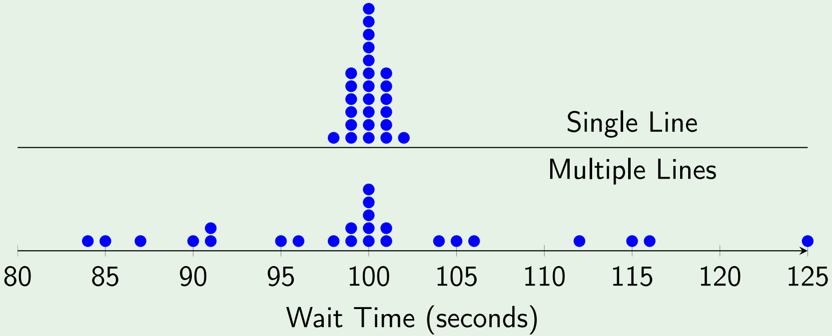
Example 1

Consider the dotplot of waiting times at a bank.



Example 1

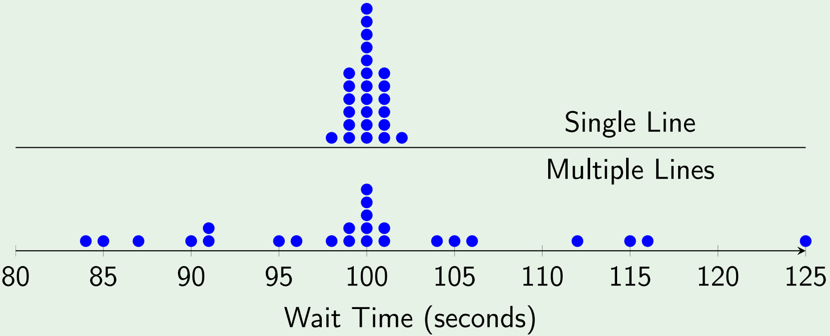
Consider the dotplot of waiting times at a bank.



Both of these data sets have the same mean, but there are clearly different.

Example 1

Consider the dotplot of waiting times at a bank.



Both of these data sets have the same mean, but there are clearly different.

The bank didn't switch to multiple lines because it made them more efficient, nor because customer wait times were reduced, but because customers prefer waiting times with less variation.

Definition

The **range** of a set of data values is the difference between the maximum data value and the minimum data value.

$$\text{Range} = (\text{maximum data value}) - (\text{minimum data value})$$

Definition

The **range** of a set of data values is the difference between the maximum data value and the minimum data value.

$$\text{Range} = (\text{maximum data value}) - (\text{minimum data value})$$

Properties

- The range is very sensitive to extreme values.

Definition

The **range** of a set of data values is the difference between the maximum data value and the minimum data value.

$$\text{Range} = (\text{maximum data value}) - (\text{minimum data value})$$

Properties

- The range is very sensitive to extreme values.
- Because the range only uses two values it does not reflect the true variation among all of the data values.

Example 2

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Example 2

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The range is

$$\text{Range} =$$

Example 2

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The range is

$$\text{Range} = 55.6 - 14.1$$

Example 2

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The range is

$$\text{Range} = 55.6 - 14.1 = 41.50 \text{ Mbps}$$

Example 2

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The range is

$$\text{Range} = 55.6 - 14.1 = 41.50 \text{ Mbps}$$

Note

Just like measures of center, we want to round to one more decimal place that our data contains.

Definition

The **standard deviation** of a set of sample values is a measure of how much data values deviate away from the mean.

Definition

The **standard deviation** of a set of sample values is a measure of how much data values deviate away from the mean.

Notation

- s is the sample standard deviation.
- σ is the population standard deviation.

Definition

The **standard deviation** of a set of sample values is a measure of how much data values deviate away from the mean.

Notation

- s is the sample standard deviation.
- σ is the population standard deviation.

Formula

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Definition

The **standard deviation** of a set of sample values is a measure of how much data values deviate away from the mean.

Notation

- s is the sample standard deviation.
- σ is the population standard deviation.

Formula

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

A shortcut version that tends to be used by software is

$$s = \sqrt{\frac{n(\sum x^2) - (\sum x)^2}{n(n - 1)}}$$

Properties

- The standard deviation is a measure of how much data values deviate from the mean.

Properties

- The standard deviation is a measure of how much data values deviate from the mean.
- The value of the standard deviation is never negative.

Properties

- The standard deviation is a measure of how much data values deviate from the mean.
- The value of the standard deviation is never negative.
- The value of the standard deviation is only zero when all of the data values are exactly the same.

Properties

- The standard deviation is a measure of how much data values deviate from the mean.
- The value of the standard deviation is never negative.
- The value of the standard deviation is only zero when all of the data values are exactly the same.
- Larger values of s indicate greater amounts of variation.

Properties

- The standard deviation is a measure of how much data values deviate from the mean.
- The value of the standard deviation is never negative.
- The value of the standard deviation is only zero when all of the data values are exactly the same.
- Larger values of s indicate greater amounts of variation.
- The standard deviation can increase dramatically with one or more outliers.

Properties

- The standard deviation is a measure of how much data values deviate from the mean.
- The value of the standard deviation is never negative.
- The value of the standard deviation is only zero when all of the data values are exactly the same.
- Larger values of s indicate greater amounts of variation.
- The standard deviation can increase dramatically with one or more outliers.
- The units of the standard deviation are the same units as the original data values.

Properties

- The standard deviation is a measure of how much data values deviate from the mean.
- The value of the standard deviation is never negative.
- The value of the standard deviation is only zero when all of the data values are exactly the same.
- Larger values of s indicate greater amounts of variation.
- The standard deviation can increase dramatically with one or more outliers.
- The units of the standard deviation are the same units as the original data values.
- The sample standard deviation is a **biased estimator** of the population standard deviation. This means that the values of s for not center around the value of σ.

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

use the following general procedure.

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

use the following general procedure.

Step 1: Compute the mean $\bar{x}$.

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

use the following general procedure.

Step 1: Compute the mean $\bar{x}$.

Step 2: Subtract the mean from each data value. (These are called deviations.)

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

use the following general procedure.

Step 1: Compute the mean $\bar{x}$.

Step 2: Subtract the mean from each data value. (These are called deviations.)

Step 3: Square each deviations from **Step 2**.

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

use the following general procedure.

Step 1: Compute the mean $\bar{x}$.

Step 2: Subtract the mean from each data value. (These are called deviations.)

Step 3: Square each deviations from **Step 2**.

Step 4: Add all the squares obtained from **Step 3**.

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

use the following general procedure.

- **Step 1:** Compute the mean $\bar{x}$.
- **Step 2:** Subtract the mean from each data value. (These are called deviations.)
- **Step 3:** Square each deviations from **Step 2**.
- **Step 4:** Add all the squares obtained from **Step 3**.
- **Step 5:** Divide the total from **Step 4** by $n - 1$.

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

use the following general procedure.

Step 1: Compute the mean $\bar{x}$.

Step 2: Subtract the mean from each data value. (These are called deviations.)

Step 3: Square each deviations from **Step 2**.

Step 4: Add all the squares obtained from **Step 3**.

Step 5: Divide the total from **Step 4** by $n - 1$.

Step 6: Find the square root of the result of **Step 5**.

Procedure

For the equation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

use the following general procedure.

Step 1: Compute the mean $\bar{x}$.

Step 2: Subtract the mean from each data value. (These are called deviations.)

Step 3: Square each deviations from **Step 2**.

Step 4: Add all the squares obtained from **Step 3**.

Step 5: Divide the total from **Step 4** by $n - 1$.

Step 6: Find the square root of the result of **Step 5**.

The result of **Step 6** is s, the standard deviation.

Example 3

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Example 3

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Step 1: $\bar{x} = \dfrac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = 30.74 \text{ Mbps}$

Example 3

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Step 1: $\bar{x} = \dfrac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = 30.74 \text{ Mbps}$

Step 2: Deviations:

7.76	24.86	-8.34	-16.64	-7.64

Example 3

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Step 1: $\bar{x} = \dfrac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = 30.74$ Mbps

Step 2: Deviations:

7.76	24.86	-8.34	-16.64	-7.64

Step 3: Squares:

60.2176	618.0196	69.5556	276.8896	58.3696

Example 3

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Step 1: $\bar{x} = \dfrac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = 30.74$ Mbps

Step 2: Deviations:

7.76	24.86	-8.34	-16.64	-7.64

Step 3: Squares:

60.2176	618.0196	69.5556	276.8896	58.3696

Step 4: Sum $= 1083.0520$

Example 3

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Step 1: $\bar{x} = \dfrac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = 30.74$ Mbps

Step 2: Deviations:

7.76	24.86	-8.34	-16.64	-7.64

Step 3: Squares:

60.2176	618.0196	69.5556	276.8896	58.3696

Step 4: Sum $= 1083.0520$

Step 5: $\dfrac{1083.0520}{4} = 270.7630$

Example 3

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Step 1: $\bar{x} = \dfrac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = 30.74$ Mbps

Step 2: Deviations:

7.76	24.86	-8.34	-16.64	-7.64

Step 3: Squares:

60.2176	618.0196	69.5556	276.8896	58.3696

Step 4: Sum $= 1083.0520$

Step 5: $\dfrac{1083.0520}{4} = 270.7630$

Step 6: $s = \sqrt{270.7630} = 16.45$ Mbps

Example 4

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Example 4

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

To use second version of the formula we need to know:

$$\sum x = 153.7$$
$$\sum x^2 = 5807.79$$

Example 4

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

To use second version of the formula we need to know:

$$\sum x = 153.7$$
$$\sum x^2 = 5807.79$$

Then,

$$s = \sqrt{\frac{n(\sum x^2) - (\sum x)^2}{n(n-1)}}$$

Example 4

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

To use second version of the formula we need to know:

$$\sum x = 153.7$$
$$\sum x^2 = 5807.79$$

Then,

$$s = \sqrt{\frac{n(\sum x^2) - (\sum x)^2}{n(n-1)}} = \sqrt{\frac{5(5807.79) - (153.7)^2}{5(5-1)}}$$

Example 4

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

To use second version of the formula we need to know:

$$\sum x = 153.7$$
$$\sum x^2 = 5807.79$$

Then,

$$s = \sqrt{\frac{n(\sum x^2) - (\sum x)^2}{n(n-1)}} = \sqrt{\frac{5(5807.79) - (153.7)^2}{5(5-1)}} = 16.45 \text{ Mbps}$$

Example 4

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), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

To use second version of the formula we need to know:

$$\sum x = 153.7$$
$$\sum x^2 = 5807.79$$

Then,

$$s = \sqrt{\frac{n(\sum x^2) - (\sum x)^2}{n(n-1)}} = \sqrt{\frac{5(5807.79) - (153.7)^2}{5(5-1)}} = 16.45 \text{ Mbps}$$

This is the same result we found in Example 3.

Range Rule of Thumb for Identifying Significant Values

A crude, but simple tool for interpreting the standard deviation.

Significantly low values are $\mu - 2\sigma$ or lower.

Significantly high values are $\mu + 2\sigma$ or higher.

Values not significant when between $\mu - 2\sigma$ and $\mu + 2\sigma$.

Range Rule of Thumb for Identifying Significant Values

A crude, but simple tool for interpreting the standard deviation.

Significantly low values are $\mu - 2\sigma$ or lower.

Significantly high values are $\mu + 2\sigma$ or higher.

Values not significant when between $\mu - 2\sigma$ and $\mu + 2\sigma$.

This works because for most data sets the vast majority of values fall within two standard deviations of the mean.

Range Rule of Thumb for Identifying Significant Values

A crude, but simple tool for interpreting the standard deviation.

Significantly low values are $\mu - 2\sigma$ or lower.

Significantly high values are $\mu + 2\sigma$ or higher.

Values not significant when between $\mu - 2\sigma$ and $\mu + 2\sigma$.

This works because for most data sets the vast majority of values fall within two standard deviations of the mean.

Range Estimate of the Standard Deviation

A rough estimate of the standard deviation is

$$s \approx \frac{\text{range}}{4}$$

Example 5

Using the full data set of Verizon data speeds (Data Set 32 Appendix B)

$$\bar{x} = 17.60 \text{ Mbps}$$
$$s = 16.02 \text{ Mbps}$$

Example 5

Using the full data set of Verizon data speeds (Data Set 32 Appendix B)

$$\bar{x} = 17.60 \text{ Mbps}$$
$$s = 16.02 \text{ Mbps}$$

Significantly low values are $(17.60 - 2 \cdot 16.02) = -14.44$ Mbps for lower.

Example 5

Using the full data set of Verizon data speeds (Data Set 32 Appendix B)

$$\bar{x} = 17.60 \text{ Mbps}$$
$$s = 16.02 \text{ Mbps}$$

Significantly low values are $(17.60 - 2 \cdot 16.02) = -14.44$ Mbps for lower.

Significantly high values are $(17.60 + 2 \cdot 16.02) = 49.64$ Mbps for lower.

Example 5

Using the full data set of Verizon data speeds (Data Set 32 Appendix B)

$$\bar{x} = 17.60 \text{ Mbps}$$
$$s = 16.02 \text{ Mbps}$$

Significantly low values are $(17.60 - 2 \cdot 16.02) = -14.44$ Mbps for lower.

Significantly high values are $(17.60 + 2 \cdot 16.02) = 49.64$ Mbps for lower.

Values that are between -14.44 Mbps and 49.64 Mbps are not significant.

Example 5

Using the full data set of Verizon data speeds (Data Set 32 Appendix B)

$$\bar{x} = 17.60 \text{ Mbps}$$
$$s = 16.02 \text{ Mbps}$$

Significantly low values are $(17.60 - 2 \cdot 16.02) = -14.44$ Mbps for lower.

Significantly high values are $(17.60 + 2 \cdot 16.02) = 49.64$ Mbps for lower.

Values that are between -14.44 Mbps and 49.64 Mbps are not significant.

Note

Based on this these results, we expect that typical airport Verizon data speeds are between -14.44 Mbps and 49.64 Mbps.

Standard Deviation of a Population

A population of size N has standard deviation

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

Standard Deviation of a Population

A population of size N has standard deviation

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

Note

When using a calculator or computer, make sure you use proper button for either the sample or population standard deviation.

Definition

The **variance** of a set of values is a measure of variation equal to the square of the standard deviation.

Definition

The **variance** of a set of values is a measure of variation equal to the square of the standard deviation.

Notation

- The sample variance is s^2.
- The population variance is σ^2.

Definition

The **variance** of a set of values is a measure of variation equal to the square of the standard deviation.

Notation

- The sample variance is s^2.
- The population variance is σ^2.

Properties

- The units of the variance are the squares of the original units.

Definition

The **variance** of a set of values is a measure of variation equal to the square of the standard deviation.

Notation

- The sample variance is s^2.
- The population variance is σ^2.

Properties

- The units of the variance are the squares of the original units.
- The value of the variance can increase dramatically with the inclusion of outliers.

Definition

The **variance** of a set of values is a measure of variation equal to the square of the standard deviation.

Notation

- The sample variance is s^2.
- The population variance is σ^2.

Properties

- The units of the variance are the squares of the original units.
- The value of the variance can increase dramatically with the inclusion of outliers.
- The value of the variance is never negative.

Definition

The **variance** of a set of values is a measure of variation equal to the square of the standard deviation.

Notation

- The sample variance is s^2.
- The population variance is σ^2.

Properties

- The units of the variance are the squares of the original units.
- The value of the variance can increase dramatically with the inclusion of outliers.
- The value of the variance is never negative.
- The value of the variance is zero only when all the data values are the same.

Definition

The **variance** of a set of values is a measure of variation equal to the square of the standard deviation.

Notation

- The sample variance is s^2.
- The population variance is σ^2.

Properties

- The units of the variance are the squares of the original units.
- The value of the variance can increase dramatically with the inclusion of outliers.
- The value of the variance is never negative.
- The value of the variance is zero only when all the data values are the same.
- The sample variance is a **unbiased estimator** of the population variance

Biased and Unbiased Estimators

- The sample standard deviation s is a **biased estimator** of the population standard deviation σ, which means that the values of the sample standard deviation do not tend to center around the value of the population standard deviation σ.

Biased and Unbiased Estimators

- The sample standard deviation s is a **biased estimator** of the population standard deviation σ, which means that the values of the sample standard deviation do not tend to center around the value of the population standard deviation σ.
  - Why individual values of s could equal for exceed σ, values of s tend to underestimate the value of σ.

Biased and Unbiased Estimators

- The sample standard deviation s is a **biased estimator** of the population standard deviation σ, which means that the values of the sample standard deviation do not tend to center around the value of the population standard deviation σ.
  - Why individual values of s could equal for exceed σ, values of s tend to underestimate the value of σ.
- The sample variance s^2 is an **unbiased estimator** of the population variance σ^2, which means that values of s^2 tend to center around the value of σ^2 instead of overestimating or underestimating.