

# Introduction to Numerical Statistics: Average and Spread

## Introduction to Numerical Statistics

Colm Mulcahy

Math 107-03, Spring 2020, Spelman College

1 Apr 2020

We discuss different measures of average and spread

## We discuss different measures of average and spread

The word “average” (in reference to a collection of numbers) is ambiguous in English and has three totally different meanings.

## We discuss different measures of average and spread

The word “average” (in reference to a collection of numbers) is ambiguous in English and has three totally different meanings.

It can refer to **mean**, **median**, or **mode**, each of which will be discussed.

## We discuss different measures of average and spread

The word “average” (in reference to a collection of numbers) is ambiguous in English and has three totally different meanings.

It can refer to **mean**, **median**, or **mode**, each of which will be discussed.

There are also several ways to measure the spread of a collection of numbers.

## We discuss different measures of average and spread

The word “average” (in reference to a collection of numbers) is ambiguous in English and has three totally different meanings.

It can refer to **mean**, **median**, or **mode**, each of which will be discussed.

There are also several ways to measure the spread of a collection of numbers.

The most important of these is the **standard deviation**: it's messy to work out especially if the **mean** (or the original collection of numbers) involves many decimals.

## We discuss different measures of average and spread

The word “average” (in reference to a collection of numbers) is ambiguous in English and has three totally different meanings.

It can refer to **mean**, **median**, or **mode**, each of which will be discussed.

There are also several ways to measure the spread of a collection of numbers.

The most important of these is the **standard deviation**: it's messy to work out especially if the **mean** (or the original collection of numbers) involves many decimals.

Today we will learn how to compute it; in future classes we will learn what it signifies and how to use it to answer interesting questions.

## Statistics: Averages (6A in text)



## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 70, 80, 90, 90, 100

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 70, 80, 90, 90, 100

The **mean** (add them up and divide by 6) is roughly 81.66666667

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 70, 80, 90, 90, 100

The **mean** (add them up and divide by 6) is roughly 81.66666667

The **median** (take the middle one, or the mean of the two middle ones) is 85.

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 70, 80, 90, 90, 100

The **mean** (add them up and divide by 6) is roughly 81.66666667

The **median** (take the middle one, or the mean of the two middle ones) is 85.

The **mode** (“the most popular one”) is 90.

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 70, 80, 90, 90, 100

The **mean** (add them up and divide by 6) is roughly 81.66666667

The **median** (take the middle one, or the mean of the two middle ones) is 85.

The **mode** (“the most popular one”) is 90.

Repeat for:

60, 70, 80, 90, 90

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 70, 80, 90, 90, 100

The **mean** (add them up and divide by 6) is roughly 81.66666667

The **median** (take the middle one, or the mean of the two middle ones) is 85.

The **mode** (“the most popular one”) is 90.

Repeat for:

60, 70, 80, 90, 90

The mean is 78, the median is 80 and the mode is 90.

## Statistics: Averages (6A in text)

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 65, 70, 75, 80, 90, 100



## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 65, 70, 75, 80, 90, 100

The mean is roughly 77.142857143, the median is 75 and there is no mode!

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 65, 70, 75, 80, 90, 100

The mean is roughly 77.142857143, the median is 75 and there is no mode!

Repeat for:

65, 65, 85, 70, 75

## Statistics: Averages (6A in text)

What is the average of these test scores?

60, 65, 70, 75, 80, 90, 100

The mean is roughly 77.142857143, the median is 75 and there is no mode!

Repeat for:

65, 65, 85, 70, 75

The mean is 72, the median is 70 and the mode is 65.

# The Mean Formula

Given a list of  $n$  numbers  $x_1, x_2, \dots, x_n$  we can compute their mean using the formula:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

# The Mean Formula

Given a list of  $n$  numbers  $x_1, x_2, \dots, x_n$  we can compute their mean using the formula:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Ugly decimals may well arise, depending on the denominator.

# The Mean Formula

Given a list of  $n$  numbers  $x_1, x_2, \dots, x_n$  we can compute their mean using the formula:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Ugly decimals may well arise, depending on the denominator.

While we can report the answer correct to 2 decimal places in many situations, we will still need to use 8 or 9 decimal places when using the mean to work out another important “summary” number, the standard deviation.

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.



# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.
2. Next, we add them up and compute the mean  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$ .

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.
2. Next, we add them up and compute the mean  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$ .
3. Then we work out the deviations: subtracting  $\bar{x}$  from each  $x_i$ .

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.
2. Next, we add them up and compute the mean  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$ .
3. Then we work out the deviations: subtracting  $\bar{x}$  from each  $x_i$ . This always gives some negative and some positive numbers.

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.
2. Next, we add them up and compute the mean  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$ .
3. Then we work out the deviations: subtracting  $\bar{x}$  from each  $x_i$ . This always gives some negative and some positive numbers. We put those in the second column of the table.

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.
2. Next, we add them up and compute the mean  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$ .
3. Then we work out the deviations: subtracting  $\bar{x}$  from each  $x_i$ . This always gives some negative and some positive numbers. We put those in the second column of the table.
5. We now work out the squares of the deviations

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.
2. Next, we add them up and compute the mean  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$ .
3. Then we work out the deviations: subtracting  $\bar{x}$  from each  $x_i$ . This always gives some negative and some positive numbers. We put those in the second column of the table.
5. We now work out the squares of the deviations: this always gives positive numbers (or zero), which we place in the third column of the table.

# Standard Deviation

The **Standard Deviation** of a collection of numbers is a number obtained from them by a multi-stage process. It is usually calculator intensive, and it's good to organize our work in a three-column table.

1. First we list the numbers vertically, in the left column of a table.
2. Next, we add them up and compute the mean  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$ .
3. Then we work out the deviations: subtracting  $\bar{x}$  from each  $x_i$ . This always gives some negative and some positive numbers. We put those in the second column of the table.
5. We now work out the squares of the deviations: this always gives positive numbers (or zero), which we place in the third column of the table. *Make sure there are no negative numbers here!*

# Standard Deviation

6. Now we add up the squares of the deviations



# Standard Deviation

6. Now we add up the squares of the deviations, write that at the bottom of the third column of the table.

# Standard Deviation

6. Now we add up the squares of the deviations, write that at the bottom of the third column of the table.
7. Divide that by  $n - 1$ .

# Standard Deviation

6. Now we add up the squares of the deviations, write that at the bottom of the third column of the table.
7. Divide that by  $n - 1$ .
8. Take the square root! WE'RE DONE!

# Overview

The 3 numbers 40, 75, 95 have mean 70.

## Overview

The 3 numbers 40, 75, 95 have mean 70.

data	data - mean	$(\text{data} - \text{mean})^2$
40	-30	900
75	5	25
95	25	625
210	-	1550

Divide 1550 by 2 to get 775.

## Overview

The 3 numbers 40, 75, 95 have mean 70.

data	data - mean	$(\text{data} - \text{mean})^2$
40	-30	900
75	5	25
95	25	625
210	-	1550

Divide 1550 by 2 to get 775.

Sqrt(775) is . . . .

## Overview

The 3 numbers 40, 75, 95 have mean 70.

data	data - mean	$(\text{data} - \text{mean})^2$
40	-30	900
75	5	25
95	25	625
210	-	1550

Divide 1550 by 2 to get 775.

Sqrt(775) is . . . .

Get about 27.83882181415