

What is Statistics, Anyway?

- Statistics is a branch of math that deals with the collection, analysis, and interpretation of information.
 - This information is called **data**.
- A characteristic of someone or something that can be classified, counted or measured is called a **variable**.

Population vs. Sample

- World population
- Entire production of chocolate chunk brownies
- All students in a school
- Population of a U.S. state
- 45 brownies from the production
- Students interviewed in English class third hour

Using this information, write your own definitions for **population** and **sample**.



Measures of Center

Using mean, median, mode to analyze data

Mode

- The value occurring most often in a set of data
- Mode = most often
- Very useful when analyzing categorical data
 - Favorite lunch item at LHS
 - The most frequently sold size of hoodie (for reordering purposes)

Mean (Arithmetic Average)

- Symbols:
 - μ - Greek letter (me-eww)
 - \bar{x}
- Add all data points and divide by the number you added
- Easily influenced by **outliers**

Median

- This is the middle of the data set
- Data must be arranged in **ascending** order
- Not easily influenced by outliers
- If set has an odd number of values, find the middle for the median.
- If set has an even number of values, find the mean of the two middle values.

Measures of center describe the typical value of a set of data

- Mean and median are considered the only “measures of center”
- Mode is not always a true typical value
- Mean is used as measure of center if the data is close together
- Median is used as measure of center if the data is spread far apart

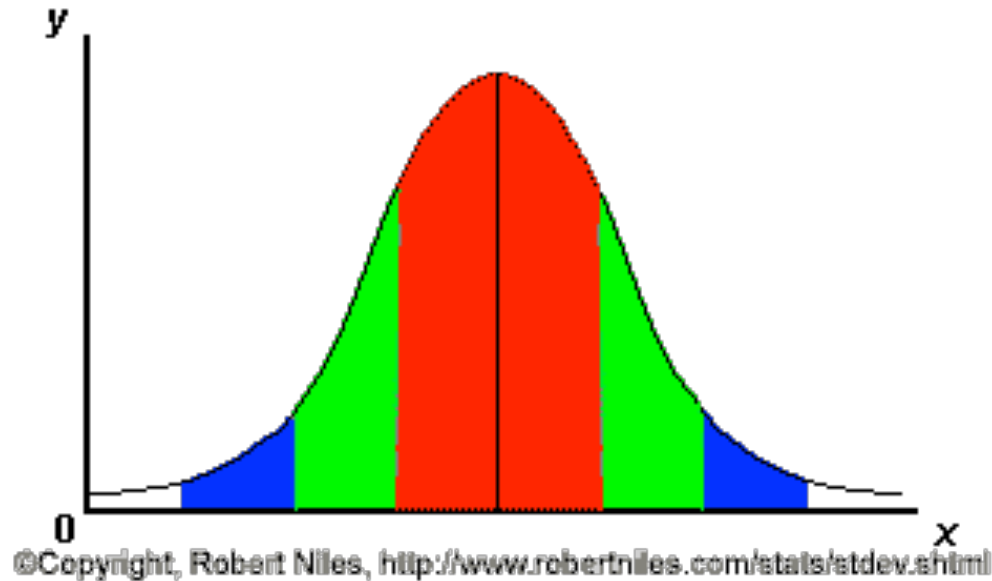
Deadliest Hurricanes in U.S.

Year	Location	# of Deaths
1900	Gavelston, TX	Approx 8,000
1928	Lake Okeechobee, FL	Approx 2,500
2005	Katrina (LA/MS)	1,800
1957	Audrey (TX/LA)	419
1935	Florida Keys	408
1926	FL/MS/AL	372
1909	LA	350
1919	Florida Keys/TX	287
1915	Two hurricanes in same season	Both were 275
1938	Great New England Hurricane	256
1969	Camille (AL/LA/FL/TX/VA/WV)	256

Source: Farmer's Almanac

Range

- The difference between the highest and lowest data values
- Low range = data is close together
- High range = data is spread apart



68% of the data fits within one standard deviation of the mean (1σ) - the red

95% of the data fits within two standard deviations of the mean (2σ) - the red and green

99% of the data fits within three standard deviations of the mean (3σ) - the red, green, and blue

Standard Deviation

- Simply put, standard deviation measures how far data values are distributed away from the mean
- It is a number
- We refer to 1 sd, 2 sd, and 3 sd
- Use your calculator to find this