Measures of Central Tendency

MATH 130, *Elements of Statistics I*

J Robert Buchanan

Department of Mathematics

Fall 2023

Introduction

Measures of central tendency are designed to provide one number which is typical of all the data values of a variable.

- mean or average
- median
- mode
- midrange

Introduction

Measures of central tendency are designed to provide one number which is typical of all the data values of a variable.

- mean or average
- median
- mode
- midrange

Definition

A **parameter** is a descriptive measure of a population.

A **statistic** is a descriptive measure of a sample.

The Mean

Definition

The **mean** of a variable is computed by the sum of all of the values of the variable in the data set divided by the number of observations. The **population mean** is denoted μ (and is a parameter). The **sample mean** is denoted $\overline{x}$ (and is a statistic).

Formulas for the Mean

Definition

If $\{x_1, x_2, \ldots, x_N\}$ are the N observations from a population, then

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum x_i}{N}.$$

If $\{x_1, x_2, \ldots, x_n\}$ are the n observations from a sample, then

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}.$$

We will generally round the number calculated for the mean to one more decimal place than found in the raw data.

Example

The following numbers represent the weight of orange M&M's measured in grams. Find the mean of this sample.

0.903
0.920
0.861
1.009
0.971
0.898
0.942
0.897

Example

The following numbers represent the weight of orange M&M's measured in grams. Find the mean of this sample.

0.903
0.920
0.861
1.009
0.971
0.898
0.942
0.897

$$\bar{x} = \frac{\sum x}{8} = \frac{7.401}{8} = 0.9251$$

Example

Consider the data in the table below which represent the lives of 40 automobile batteries measured to the nearest tenth of a year.

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.8	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

1. Treating the last row of data as a sample, find the sample mean.

2. Find the population mean.

Example

Consider the data in the table below which represent the lives of 40 automobile batteries measured to the nearest tenth of a year.

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.8	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

1. Treating the last row of data as a sample, find the sample mean.

$$\bar{x} = \frac{28.6}{8} = 3.58$$

2. Find the population mean.

Example

Consider the data in the table below which represent the lives of 40 automobile batteries measured to the nearest tenth of a year.

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.8	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

1. Treating the last row of data as a sample, find the sample mean.

$$\bar{x} = \frac{28.6}{8} = 3.58$$

2. Find the population mean.

$$\mu = \frac{\sum x}{N} = \frac{136.4}{40} = 3.41$$

The Median

Definition

The **median** of a variable is the value that lies in the middle of the data when arranged in ascending order. Half the data are below the median and half are above the median. The median is denoted M.

Note: you may also see the median denoted as $\tilde{x}$.

How to Find the Median

1. Arrange the data in ascending order.
2. Determine the number of observations and call it n.
3. If n is odd, the median is the observation in the $\left(\frac{n+1}{2}\right)$ position. If n is even, the median is the mean of the two middle observations.

Example

The following numbers represent the weight of green M&M's measured in grams.

0.890
0.902
0.902
0.911
0.930
0.949
1.002

Find the median weight.

Example

The following numbers represent the weight of green M&M's measured in grams.

0.890
0.902
0.902
0.911
0.930
0.949
1.002

Find the median weight.
Since the data are already in ascending order and $n = 7$, we need the $(7 + 1)/2 = 4$th value,

$$M = 0.911.$$

Example

Consider the automobile battery lifespan data from a previous example.

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.8	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

Find the median lifespan.

Example

Consider the automobile battery lifespan data from a previous example.

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.8	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

Find the median lifespan.
The data are in ascending order arranged by columns and $n = 40$, thus the median is the average of the 20th and 21st values.

$$M = \frac{3.4 + 3.4}{2} = 3.40$$

Identifying the Shape of a Distribution

Distribution Shape	**Mean *vs.* Median**
Symmetric	Mean nearly equal to median
Skewed left	Mean smaller than the median
Skewed right	Mean larger than the median

Identifying the Shape of a Distribution

Distribution Shape	**Mean *vs.* Median**
Symmetric	Mean nearly equal to median
Skewed left	Mean smaller than the median
Skewed right	Mean larger than the median

Definition

A numerical summary of data is said to be **resistant** if extreme values (very large or very small) relative to the data do not affect its value substantially.

Identifying the Shape of a Distribution

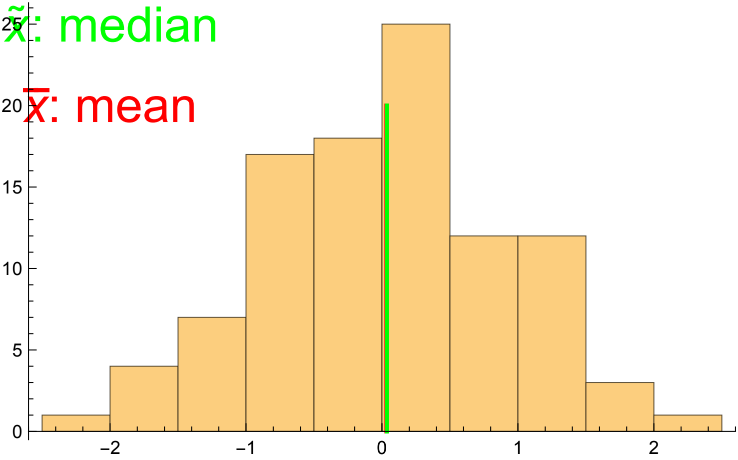
Distribution Shape	Mean *vs.* Median
Symmetric	Mean nearly equal to median
Skewed left	Mean smaller than the median
Skewed right	Mean larger than the median

Definition

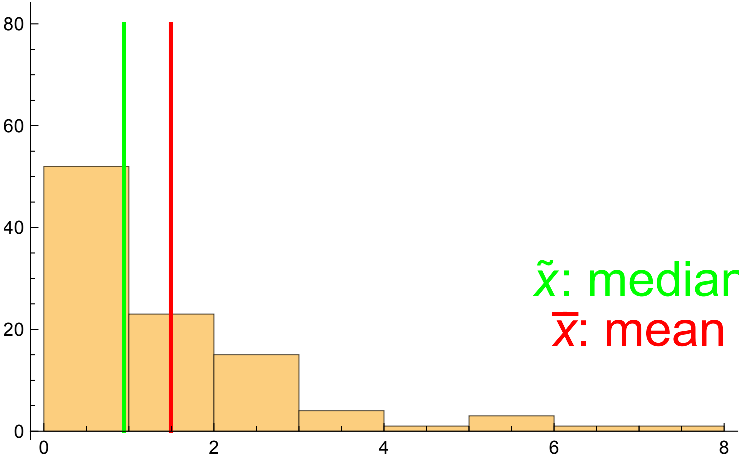
A numerical summary of data is said to be **resistant** if extreme values (very large or very small) relative to the data do not affect its value substantially.

Remark: the median is resistant, the mean is not resistant.

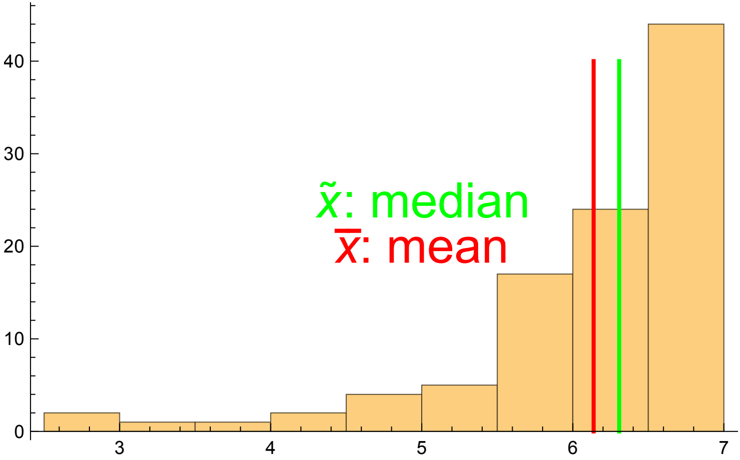
Example: Symmetric



Example: Skewed Right



Example: Skewed Left



The Mode

Definition

The **mode** of a variable is the most frequent observation that occurs in the data set.

- To find the mode, tally the number of observations for each value of the variable.
- If no observation occurs more than once, the data have no mode.
- There can be more than one mode.

Example

Find the mode of the lifespans of the automobile batteries.

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.8	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

Solution

Summarized as a stem-and-leaf plot we see the following distribution of data.

1.	6 8
2.	2 5 6 6 9
3.	0 0 1 1 1 1 2 2 2 3 3 3 4 4 4 5 5 6 7 7 7 8 8 9 9
4.	1 1 2 3 4 5 7 7

Inspecting this stem-and-leaf plot carefully we see that the

$$\text{mode} = 3.1.$$

Example

Find the mode of the weights of the following blue M&M's.

0.890
0.902
0.902
0.911
0.930
0.949
0.949
1.002

Example

Find the mode of the weights of the following blue M&M's.

0.890
0.902
0.902
0.911
0.930
0.949
0.949
1.002

There are two modes (in this case we can say the data are **bimodal**).

$$\text{mode} = 0.902 \text{ and } 0.949$$

The Midrange

Definition

The **midrange** of a variable is the average of the lowest and highest values of the variable found in the data set.

Example

Find the midrange of the set of weights of red M&M's.

0.870	0.933
0.952	0.908
0.911	0.908
0.913	0.983
0.920	0.936
0.891	0.924
0.897	0.908
0.924	0.897
0.912	0.888
0.898	0.882

Example

Find the midrange of the set of weights of red M&M's.

0.870	0.933
0.952	0.908
0.911	0.908
0.913	0.983
0.920	0.936
0.891	0.924
0.897	0.908
0.924	0.897
0.912	0.888
0.898	0.882

The lowest weight is 0.870 and the highest weight is 0.983, thus the

$$\text{midrange} = \frac{0.870 + 0.983}{2} = 0.9265.$$

Pros and Cons of the Mean, Median, and Mode

- The most appropriate measure of central tendency will depend on the data.

Pros and Cons of the Mean, Median, and Mode

- The most appropriate measure of central tendency will depend on the data.
- The mode can be used for both qualitative and quantitative data.

Pros and Cons of the Mean, Median, and Mode

- The most appropriate measure of central tendency will depend on the data.
- The mode can be used for both qualitative and quantitative data.
- For small data sets (relatively few observations) the mean is influenced by extreme values, but the median is resistant.

Pros and Cons of the Mean, Median, and Mode

- The most appropriate measure of central tendency will depend on the data.
- The mode can be used for both qualitative and quantitative data.
- For small data sets (relatively few observations) the mean is influenced by extreme values, but the median is resistant.
- For large data sets (many observations) the mean and median tend to be close to each other.

Pros and Cons of the Mean, Median, and Mode

- The most appropriate measure of central tendency will depend on the data.
- The mode can be used for both qualitative and quantitative data.
- For small data sets (relatively few observations) the mean is influenced by extreme values, but the median is resistant.
- For large data sets (many observations) the mean and median tend to be close to each other.
- The mean is easier to calculate than the median since we do not have to sort the data.