

Measure of central tendency

A **measure of central tendency** tells us the middle of some numbers.

Example: If three friends have 2, 4, and 6 candies, the middle (average) is **4 candies**.

The main purpose of computing measures of central tendency is to give you an idea of what a typical or common value for a given variable is. The three most common measures of central tendency are the

- arithmetic mean,
- the median, and
- the mode.

The Mean

- The arithmetic mean, or simply the mean, is often referred to in ordinary speech as the average of a set of values.
- Calculating the mean as a measure of central tendency is appropriate for interval and ratio data

The mean of a population is denoted by the Greek letter mu (μ) whereas the mean of a sample is typically denoted by a bar over the variable symbol: for instance, the mean of x would be written and pronounced “ x -bar.” Or \bar{X} .

- Suppose we have a population with only five cases, and these are the values for members of that population for the variable x : 100, 115, 93, 102, 97.

- $\mu = (100 + 115 + 93 + 102 + 97)/5 = 507/5 = 101.4$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

- Consider one simple example. Suppose the last value in our tiny data set was 297 instead of 97. In this case, the mean would be:

$$\mu = (100 + 115 + 93 + 102 + 97)/5 = 507/5 = 101.4$$

- The mean of 141.4 is not a typical value for this data, In fact, 80% of the data (four of the five values) are below the mean, which is distorted by the presence of one extremely high value.
- The mean can also be calculated using data from a frequency table

Table 4-1. Simple frequency table

Value	Frequency
1	7
2	5
3	12
4	2

$$\mu = \frac{(1 \times 7) + (2 \times 5) + (3 \times 12) + (4 \times 2)}{(7 + 5 + 12 + 2)} = 2.35$$

The Median

- The median of a data set is the middle value when the values are ranked in ascending or descending order.
- If there are n values, the median is formally defined as the $(n + 1)/2$ th value, so if $n = 7$, the middle value is the $(7+1)/2$ th or fourth value.
- If there is an even number of values, the median is the average of the two middle values.
- This is formally defined as the average of the $(n / 2)$ th and $((n / 2) + 1)$ th value.
- If there are six values, the median is the average of the $(6/2)$ th and $((6/2) + 1)$ th value, or the third and fourth values.

Example

- Odd number (5) of values: 1, 4, 6, 6, 10;

Median = 6 because $(5+1)/2 = 3$, and 6 is the third value in the ordered list.

- Even number (6) of values: 1, 3, 5, 6, 10, 15;

Median = $(5+6)/2 = 5.5$ because $6/2 = 3$ and $[(6/2) + 1] = 4$, and 5 and 6 are the third and fourth values in the ordered list.

- The median is a better measure of central tendency than the mean for data that is asymmetrical or contains outliers.
- This is because the median is based on the ranks of data points rather than their actual values, and by definition, half of the data values in a distribution lie below the median and half above the median, without regard to the actual values in question.

For instance, the median of all three of the following distributions is 4:

- Distribution A: 1, 1, 3, 4, 5, 6, 7
- Distribution B: 0.01, 3, 3, 4, 5, 5, 5
- Distribution C: 1, 1, 2, 4, 5, 100, 2000

The median seems reasonably representative of the data values in Distributions A and B, but perhaps not for Distribution C.

Mode

A third common measure of central tendency is the mode, which refers to the most frequently occurring value.

The mode is most often useful in describing ordinal or categorical data.

For instance, imagine that the following numbers reflect the favored news sources of a group of college students, where 1 = newspapers, 2 = television, and 3 = Internet:

1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3

We can see that the Internet is the most popular source because 3 is the modal (most common) value in this data set.