

Measures of Central Tendency

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics.

Each of these measures of central tendency has advantages and disadvantages, the logic underlying the calculation procedure for each measure is different.

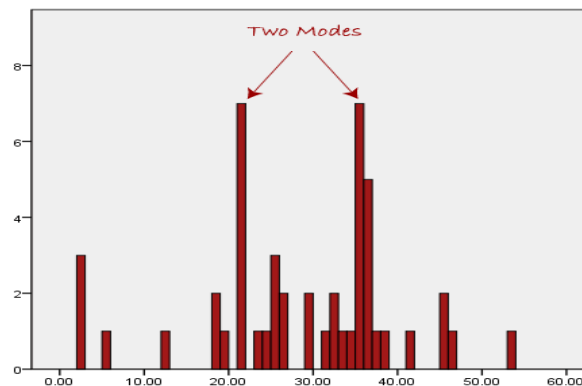
The discussion in this section will be limited to three common measures of central tendency; mode, median, and arithmetic mean.

The mode

Mode is defined as the value that occurs most frequently in the data.

Some data sets do not have a mode because each value occurs only once.

On the other hand, some data sets can have more than one mode. This happens when the data set has two or more values of equal frequency which is greater than that of any other value.

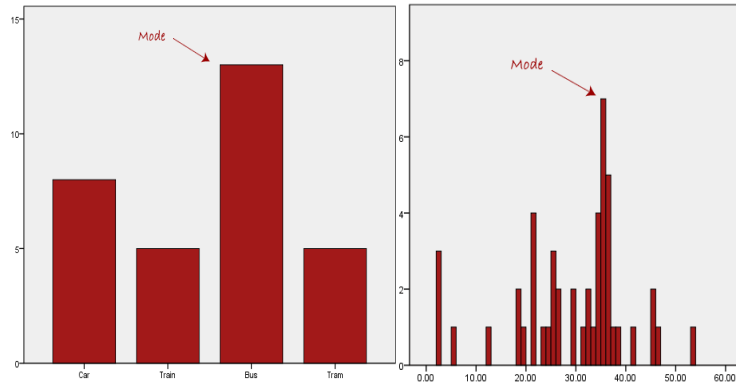


We are now stuck as to which mode best describes the central tendency of the data. This is particularly problematic when we have continuous data because we are more likely not to have any one value that is more frequent than the other.

Mode is rarely used as a summary statistic except to describe a bimodal distribution. In a bimodal distribution, the taller peak is called the major mode and the shorter one is the minor mode.

When nominal data are used, the mode would be the category containing the largest number of observations. With ordinal or interval-ratio data grouped into classes, the category with the largest number of observations is defined as the modal class.

On a histogram it represents the highest bar in a bar chart or histogram. You can, therefore, sometimes consider the mode as being the most popular option. An example of a mode is presented below:



Normally, the mode is used for categorical data where we wish to know which is the most common category, as illustrated before.

With nominal data, the mode is the only available descriptive measure of central tendency.

Although a useful measure of central tendency in many data sets, the mode may not always provide a practical result.

Worked Example:

Find the mode of this set of data:

A (5, 6, 2, 4, 7, 8, 3, 5, 6, 6)

B (6, 7, 7, 3, 8, 3, 9, 5)

Solution:

A) From the list above it is easy to see that 6 appears more than any other number, so

Mode = 6

B) Mode = 3, 7 (bimodal)

For grouped data, the mode corresponds to the top of the frequency curve.

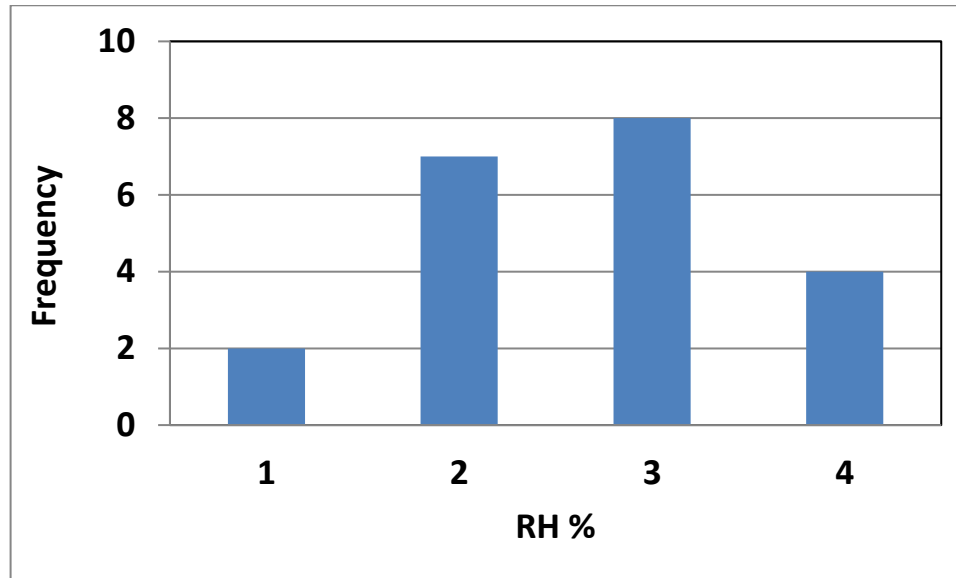
Estimating the Mode from Grouped Data

Again, looking at our data:

Seconds	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4

We can easily find the modal group (the group with the highest frequency), which is 61 - 65

We can say "the modal group is 61 - 65".



MEDIAN

Median is the value which occupies the middle position when all the observations are arranged in an ascending/descending order. It divides the frequency distribution exactly into two halves. Fifty percent of observations in a distribution have scores at or below the median. Hence median is the 50th percentile. [2] Median is also known as ‘positional average’. It is easy to calculate the median.

If the number of observations are odd (in the ordered set):

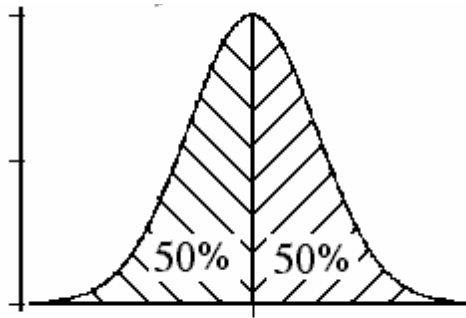
$$\text{Median} = \frac{n + 1}{2}$$

But when the total number of observations are even, it is given by (in the ordered set) :

$$\text{Median} = \frac{\text{Med1} + \text{Med2}}{2}$$

$$\text{Med1} = \frac{n+1}{2} \quad \text{and} \quad \text{Med2} = \frac{n}{2} + 1$$

For group data, the median line halves the area under the frequency curve.



Median

Worked Example

Five people play golf and at one hole their scores are: 3, 4, 4, 5, 7 For these scores, find the median

Solution

The numbers are already in order and the middle number is 4.
So median 4

Worked Example

Find the median of this set of data: (5, 6, 2, 4, 7, 8, 3, 5, 6, 6)

Solution

To find the median, place all the numbers in order.

2, 3, 4, 5, 5, 6, 6, 6, 7, 8

As there are two middle numbers in this example, 5 and 6,

median $(5+6) / 2 = 5.5$

We have to find out the median score from the following frequency distribution table:

Score	Frequency	Cumulative frequency
0-10	4	4
10-20	3	7
20-30	5	12
30-40	6	18
40-50	7	25
	Total = 25	

Now, $N=25$

$$N/2=12.5,$$

which means that we have to look for the interval which covers 12th item and 13th item.

Looking at the cumulative frequencies, we see that the 3rd interval (i.e. 20-30) covers the 12th item, while 4th interval (i.e. 30-40) covers the 13th item.

If we are supposed to take both the intervals as median class for the sake of using the formula:

$$\text{Median} = L + \frac{\frac{N}{2} - C}{F} * W$$

then we will end up with two medians. We can take the average of these as the required median, though. I want to know the correct procedure here.

In the formula,

L = lower boundary of the median class

N = total frequency

C = cumulative frequency of the class preceding the median class

f = frequency of the median class

w = width of the median class i.e. upper boundary - lower boundary

We consider the interval 20-30 as the median class and use the above formula, and then the median will be

$$\text{Median} = 20 + \frac{\frac{25}{2} - 7}{5} * 10$$

$$\text{Median} = 31$$