Chapter 3: The measures of central tendency

Mean, Median and the Mode are the indices of measures of central tendency.

3.1. The arithmetic mean

The most popular and widely used measure for representing the entire data by one value is what most laymen call an "average" and what the statisticians call the arithmetic mean. Its value is obtained by adding together all the observations and by dividing this total by the number of observations.

Arithmetic Mean-Ungrouped Data.

For ungrouped data, arithmetic mean may be computed by applying any of the following methods: Direct method, and Short-cut method.

The arithmetic mean, often simply referred to as mean, does their total number of observations divide the total of the values of a set of observations. Thus, if $X_1, X_2, \ldots\ldots s \ldots\ldots X_N$, represent the values of N items or observations; the arithmetic mean denoted by $\overline{X}$ is defined as:

In the case of the simple data, X1, X2, X3, …X_n, the arithmetic mean is:

$$\overline{X} = \frac{X_1 + X_2 + X_{3+} \ldots\ldots + X_N}{N} \quad \text{or} \quad \overline{X} = \frac{1}{N}\sum_{i=1}^{n} X_i, \quad i = 1, 2, 3, \ldots\ldots, n$$

Example1:

Let given the follow number 2, 3, 4, 5, 6. Find their mean, which is:

$$\bar{x} = \frac{2+3+4+5+6}{5} = \frac{20}{5} = 4, \quad \textit{so, the mean here is } \bar{x} = 4,$$ The formula becomes: $\overline{X} = \frac{\sum X}{N}$

Example2:

Monthly income of 10 employees working in a firm is as follows: Find its average monthly income

4487 4493 4502 4446 4475 4492 4572 4516 4468 4489.

Let income be denoted by X.

$\Sigma X = 4487 + 4493 + 4502 + 4446 + 4475 + 4492 + 4572 + 4516 + 4468 + 4489 = 44940$

$\overline{X} = \frac{\sum X}{N} = \frac{44940}{10} = 4494$. Hence, the average monthly income is 4494.

Arithmetic Mean-Grouped data

Direct method, Short-cut method.

In the case when the data are grouped in class intervals, we have to consider the mid-point or mid value of each class, and the procedures are the same as the case in 2^{nd} point.

Direct Method. When direct method is used: $\overline{X} = \frac{\sum fX}{N}$

Where: X mid-point of various classes or n_i: the frequency of each class, N: the total frequency.

Feedback

To get solutions for this activity, you have to follow the conditions of direct and shortcut method, and respect their steps. The measures of central tendency will be respected and the appropriate formulas in simple and group data will be respected. Frequencies and their cumulative play an important place to determine proportional and percentage in descriptive tables.

3.2. The median

Median is the measure of central tendency, which appears in the "middle" of an ordered sequence of values. That is, half of the observations in a set of data are lower than it and half of the observations are greater than it. For example, if the income of five persons were $ 7000, 7200, 7500, 7600, 7800, then the median income would be $ 7500.

Median-Ungrouped data: case when the number of observations is odd

Arrange the data in ascending or descending order of magnitude. (Both arrangements would give the same answer.)

If **N** is an odd numbe ($\frac{N+1}{2}$) th value is the Median.

Example: Given the numbers: 4, 7, 8, 9, 12, 15, 16. Here n=7, **Solution:** the Median is the size of $\frac{N+1}{2}$ *th observation*, otherwise, median is $X_{\frac{N+1}{2}}$, here the median is $X_4 = 9$.

Median-Ungrouped data: case when the number of observations is even.

If **N** is an even number, the median is given by: $Me = \frac{1}{2}(X_{\frac{N}{2}} + X_{\frac{N}{2}} + 1)$

Example: Let given these numbers: 5 6 10 12 14 17 23 30

Solution: The median is the size of $Me = \frac{1}{2}(X_{\frac{N}{2}} + X_{\frac{N}{2}} + 1) = \frac{1}{2}(X_4 + X_5) = \frac{1}{2}(12 + 14) = 13$

Example 1:

From the following data of wages of 7 workers, compute the median wage:

Wages ($): 4600 4650 4580 4690 4660 4606 4640

Me= X_4 , is 4640, after arrangement the above number either in ascending or descending order. Value of 4th observation is 4640. Hence, median of wages is $4640. In the above example, the number of observations was odd and, therefore, it was possible to determine the value of 4th observation. When the number of observations is even, for example, if in the above case the numbers of observations are 8 the median would be the value of $\frac{8+1}{2} = 4.5$ th observation.

For finding out the value of 4.5th, $X are 50, 80, 75, 65, 85, 45, 95, and 55. We shall take the average of 4th and 5th observations, after arrangement of the numbers, either ascending or descending order. Arrangement of the number in ascending order: 45, 50, 55, 65, 75, 80, 85, 95.

Hence the median shall be: $\frac{65+75}{2} = 70$

Median – Grouped Data: PROCEDURE:

Step I : Calculate the cumulative frequency of each class;

Step II: you have N, now you deduce its half $\frac{N}{2}$ which locate the median class

StepIII: The class corresponding to the value which is just equal to $\frac{N}{2}$, Or just greater than $\frac{N}{2}$, this is the median class.

StepIV: Apply the following formula for determining the exact value of median

$$Median = L + \frac{C}{f}(N/2 - p.cf).$$

With: L: Lower limit of median class, i, e, the class in which the middle observation in the distribution lies,

p.c.f.: Preceding cumulative frequency to the Median class,

f :Frequency of the median class,

C : the class-Interval of the median clas C: is the width or length of the data grouped in class-intervals

Feedback

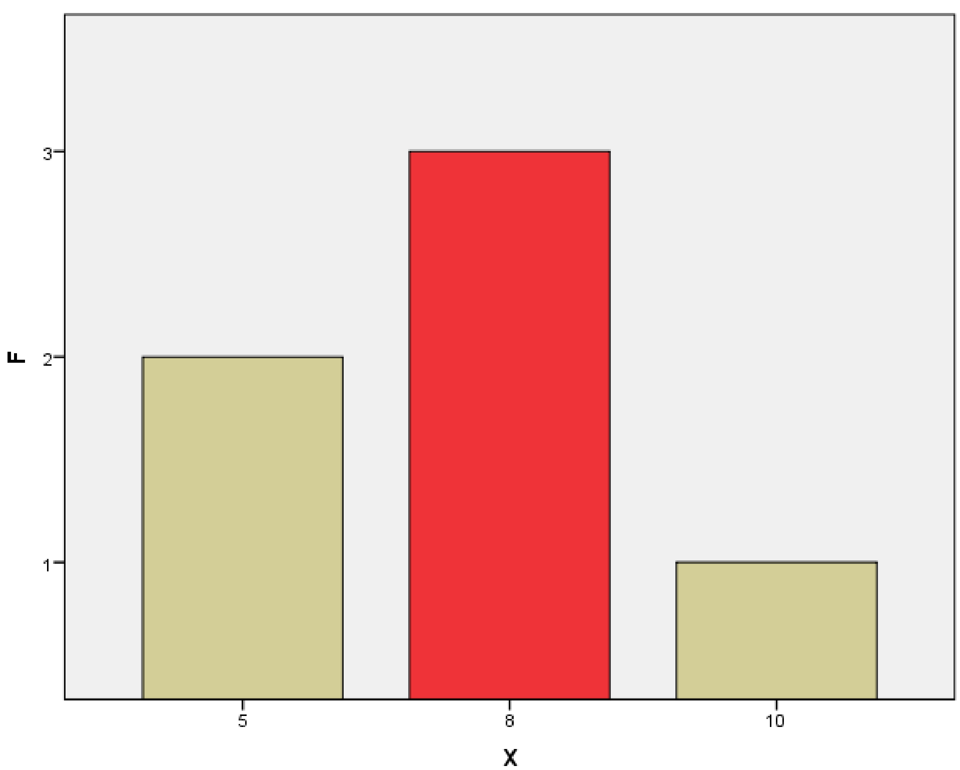
To get solutions for this activity, you have to follow the conditions of direct and shortcut method, and respect their steps. The measures of central tendency will be respected and the appropriate formulas in simple and group data will be respected. Frequencies and their cumulative play an important place to determine proportional and percentage in descriptive tables.

3.3. The mode

Mode is defined as that value which occurs more time than the others, i.e., having the maximum frequency in the given distribution.

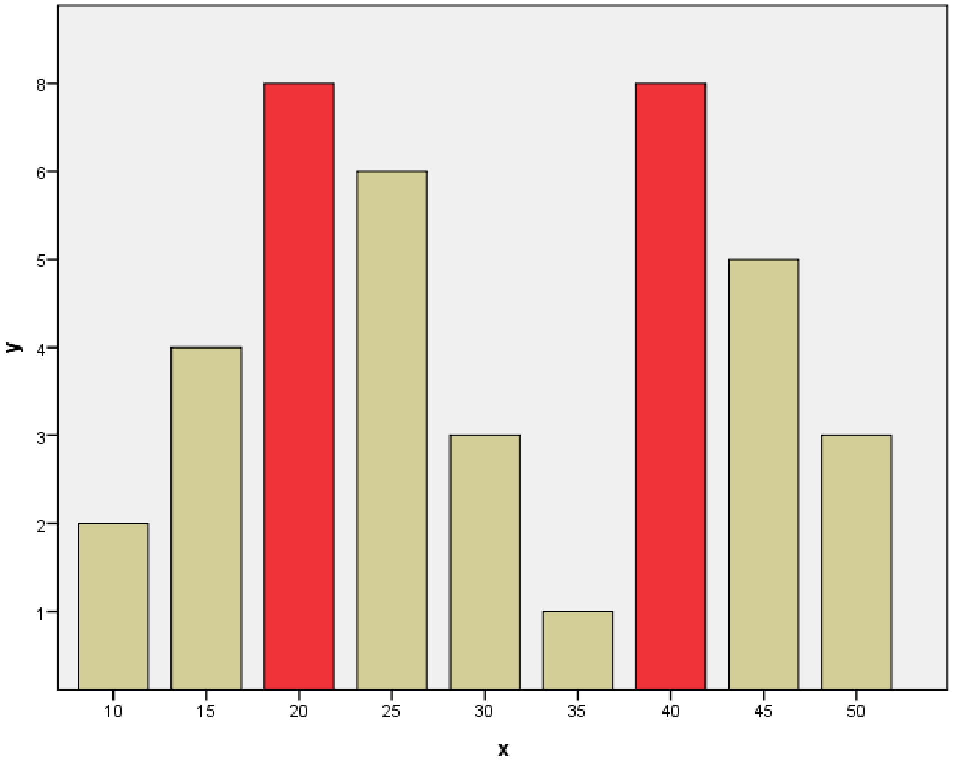
Simple data without frequencies

For example, if we take the values of six different observations as 5, 8, 10, 8, 5, 8, the mode will be 8 as it has occurred maximum number of times, i.e., 3 times. Graphically, it is the value on the X-axis below the peak, or highest point, of the frequency curve as can be seen from the following diagram.



[redacted]

In statistics, the mode only tells us which single value occurs most often; it may, therefore, represent a majority of the total population. It is possible that a distribution will be bimodal. This happens when there may be two or more values of equal or nearly equal occurrence as can be seen from the following diagram:



[redacted]

The two red colour bars show the two modes called bimodal. The presence of more than one mode has a special significance in statistical analysis, for it indicates potential trouble. It is usually dangerous to compare bimodal population or to draw conclusions about them because they usually arise when there is some non-homogeneous factor present in the population. If the collected data produce a binomial distribution, the data themselves should be questioned.

Simple data with frequencies

The mode refers to the value, which occurs most frequently in a distribution. Mode is the easiest to compute since it is the value corresponding to the maximum frequency.

For example, if the data is:

Size of shoes	:	5	6	7	8	9	10	11
No. of persons	:	10	20	25	40	22	15	6

The modal size is “8” since it appears maximum number of times in the data.

Determining the precise value of the mode of a frequency distribution is by no means an elementary calculation. Essentially, it involves fitting mathematically some appropriate type of frequency curve to the grouped data and the determination of the value on the X-axis below the peak of the curve. However, there are several elementary methods of *estimating* the mode. These methods have been discussed for ungrouped and grouped data

Mode – Ungrouped Data.

For determining mode count, the number of observations that various values repeat themselves and the value which occurs the maximum number of times is the modal value. For determining mode form the number of observations whose the value repeated and occurs mostly than others, that value is the mode.

Example:

The following figures relate to the preferences with regard to the size of screen (in inches) of T.V. sets of 30 persons selected at random from a locality. Find the modal size of the T.V. screen.

12	20	12	24	29
20	12	20	29	24
24	20	12	20	24
29	24	24	20	24
24	20	24	24	12
24	20	29	24	24

Size in inches	Tally	Frequency
12	IIIII	5
20	IIIII III	8
24	IIIII IIIII III	13
29	IIII	4
	Total: 30	

Since size 24 occurs the maximum number of times, therefore, the modal size, of T.V. screen is 24 inches.

Mode – Grouped data. $\frac{N}{2}$, In the case of grouped data, the following formula is used for calculating mode:

$$M_o = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * C$$

Where: L = Lower limit of the modal class,

$\Delta_1 =$ The difference between the frequency of the modal class and the frequency of the pre-modal class, i.e., preceding class.

$\Delta_2 =$ The difference between the frequency of the modal class and the frequency of the post-modal class, i.e., succeeding class.

C: The size of the modal class. Another form of this formula is:

$$M_O = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} x\, C$$

Where: L : Lower limit of the modal class,

f_1 : Frequency of the modal class,

f_0 : Frequency of the class preceding the modal class,

f_2 : Frequency of the class succeeding the modal class.

While applying the above formula for calculating mode, it is necessary to see that the class intervals are *uniform* throughout. If they are unequal, they should first be made equal on the assumption that the frequencies are equally distributed throughout the class, otherwise we will get misleading results. A distribution having only one is called *unimodal*. If it contains more than one mode, it is called *bimodal* or *multimodal.* In that case, the value of mode cannot be determined by the above formula and hence mode is *ill-defined* when there is more than one value of mode. A distribution with two mode is called bimodal, and more than two, it is called multimodal. When mode is ill defined, its value may be ascertained by the following approximate formula based upon the relationship between mean, median and mode. **Mode = 3 Median – 2 Mean.**