Measures of central tendency

Measures of central tendency are numerical indices that attempt to answer the question: what is the typical value of the observations in this distribution?
There are several different measures of central tendency. Each is an indicator of what a typical value is, but each employs a different definition of 'typical'.
Among the several averages, the most commonly used averages are

(i) Mean
(ii) Median and
(iii) Mode

The mean can again be of three types:
a) Arithmatic mean
b) Geometric mean and
c) Harmonic mean

Arithmatic mean:

The arithmatic mean, which is sometimes referred to as simply the mean, is the most commonly used central value of a distribution. A layman views this mean simply as an 'average'.

The arithmatic mean is the sum of a set of observations, positive, negative or zero, divided by the number of such observations.

The AM for ungrouped data:

Suppose, $x_1, x_2, \ldots, x_n$ denote the values for a variable X (n being the number of observations). Then the AM, denoted by $\bar{x}$, is defined as:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The AM for grouped data:

Suppose, for a grouped data, $x_1, x_2, \ldots, x_K$ are the mid-values of K classes and $f_1, f_2, \ldots, f_K$ are the corresponding frequencies such that $f_1 + f_2 + \cdots + f_K = n$. Then

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_K x_K}{f_1 + f_2 + \cdots + f_K} = \frac{1}{n}\sum_{i=1}^{K} f_i x_i$$

Weighted Arithmatic Mean:

The weighted mean of a set of n numbers $x_1, x_2, \ldots, x_n$ whose relative importance is measured by a corresponding set of weights $w_1, w_2, \ldots, w_n$ is given by the formula-

$$\bar{x}_w = \frac{\Sigma w_i x_i}{\Sigma w_i}$$

Example:

We may use x as a variable for age of the students in the school children.

$x_1 = 16$, $x_2 = 18$, $x_3 = 17$, $x_4 = 15$, $x_5 = 17$, $x_6 = 16$

Average, $\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{n} = \frac{16 + 18 + 17 + 15 + 17 + 16}{6}$

$=$

Example: A sample survey of Bangladesh Bureau of statistics in a rural area of Bangladesh collected the age at first marriage in years of 330 newly married women.

Age at first marriage (years)	Number of girls
11	17
12	28
13	37
14	52
15	70
16	48
17	36
18	23
19	11
20	8
Total	330

Calculate the mean age at first marriage in the sample area.

Age at marriage (x_i)	Number of girls (f_i)	Product ($f_i x_i$)
11	17	187
12	28	336
13	37	481
14	52	728
15	70	1050
16	48	768
17	36	612
18	23	414
19	11	209
20	8	160
Total	330	4945

Here, $k = 10$, $\Sigma f_i x_i = 4945$, $\Sigma f_i = 330$

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{4945}{330} = 14.98 \text{ years}.$$

Example

suppose a student in his final examination obtained 74 marks in English, 67 in mathematics, 71 in statistics, 80 in physics and 53 in chemistry. For an assessment of his final standard, if is agreed that the subjects should be weighted by the numbers 3, 4, 4, 2, 1 respectively. What is his final standard?

$$\# \quad WAM = \frac{3\times74 + 4\times67 + 4\times71 + 2\times80 + 1\times53}{3+4+4+2+1} = \frac{987}{14} = 70.5$$

Median:

The median is that value for which 50 percent of the observations, when arranged with respect to their magnitude values, either in ascending or descending order, lies on each side. Me is sometimes called positional averages.

The following steps may be followed in the computation of the median from ungrouped data;

a) List the observation in order of magnitude.

b) Count the number of observations. This is n.

c) The median is the value that corresponds to the observation number $\frac{1}{2}(n+1)$ if n is odd.

d) The median is the value that corresponds to the observation number $\frac{1}{2}\left[\frac{n}{2}+\left(\frac{n}{2}+1\right)\right]$. if n is even.

Example: The weights of 11 mothers in kg were recorded as follows:

47, 44, 42, 41, 58, 52, 55, 39, 40, 43 and 61.

Solution: To obtain the median weight, we arrange the values in ascending order. When we do so, the series becomes 39, 40, 41, 42, 43, 44, 47, 52, 55, 58 and 61. Since n is odd, the median is the value that belongs to the observation number $\frac{(n+1)}{2}$ i.e $\frac{11+1}{2} = 6$th observation. On counting, the 6th observation is 44 and hence it is the median.

Example: If one woman with weight 52 kg is removed from the study, the series becomes 39, 40, 41, 42, 43, 44, 47, 55, 58 and 61, in which case n 10, which is an even number. By definition the median will be the average of the 5th and 6th observations. This value is $\frac{(43+44)}{2} = 43.5$ which is the median.

Calculation of Median - Grouped data:

Determine the particular class in which the value of median lies. Use $\frac{N}{2}$ to locate the median class, which divides the area of the curve into two equal parts.

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

$$Me = L_0 + \frac{\left(\frac{n}{2} - F_{-Me}\right)}{f_{Me}} \times W_{Me} \quad \text{----} (*)$$

where, L_0 : lower boundary of the median class

F_{-Me} : Cumulative frequency for the class next lower to the median class (can also be called the pre-median class)

f_{Me} : frequency of the median class

W_{me} : Width of the interval of the median class.

n : Number of observations.

Example: 1500 workers are working in an industrial establishment. Their age is classified as follows:

Age (years)	No. of workers (f)	c.f
18-22	120	120
22-26	125	245
26-30	280	525
30-34	260	785
34-38	155	940
38-42	184	1124
42-46	162	1286
46-50	86	1372
50-54	75	1447
54-58	53	1500

Calculate the median age.

Median = Size of $\frac{N}{2}$ th observation = $\frac{1500}{2} = 750$ th observation

Hence median lies in the class 30-34.

$$\text{Median} = L + \frac{N/2 - f_{Me}}{f_{Me}} \times W_{Me}$$

$$= 30 + \frac{750 - 525}{260} \times 4 = 30 + 3.46$$

$$= 33.46$$

Hence the median age of the workers is 33.46 years

$$Q_i = L_i + \frac{\frac{iN}{4} - f_{-i}}{f_i}$$

Scanned with CamScanner

Mode :

Mode is the value of a distribution for which the frequency is maximum. In other words, mode is the value of a variable, which occurs with the highest frequency.

It is the simplest, but the least precise, measure of central tendency. And for nominal data, the mode is the only measure of central tendency.

For the data sets :

1. 7, 8, 6, 7, 9, 7, 4 ; the modal value is 7 and the data is unimodal.

2. 6, 5, 7, 5, 2, 3, 3 ; the modal values are 5 and 3 and the data is bimodal.

3. 1, 5, 7, 2, 6, 9, 4 ; there is no modal value

\# nominal data : the Mo

Ordinal data : the Me

Interval or Ratio data : the AM

Skew distribution : the ~~AMe~~ Me

Open-end distribution : the Me

$$M_o = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times W_{Mo}$$

where, L_1 : the lower boundary of the modal class

$\Delta_1 = f_{Mo} - f_{-Mo}$

$\Delta_2 = f_{Mo} - f_{+Mo}$

W_{Mo} : width of the modal class

f_{-Mo} : frequency of pre-modal class

f_{+Mo} : frequency of post modal class

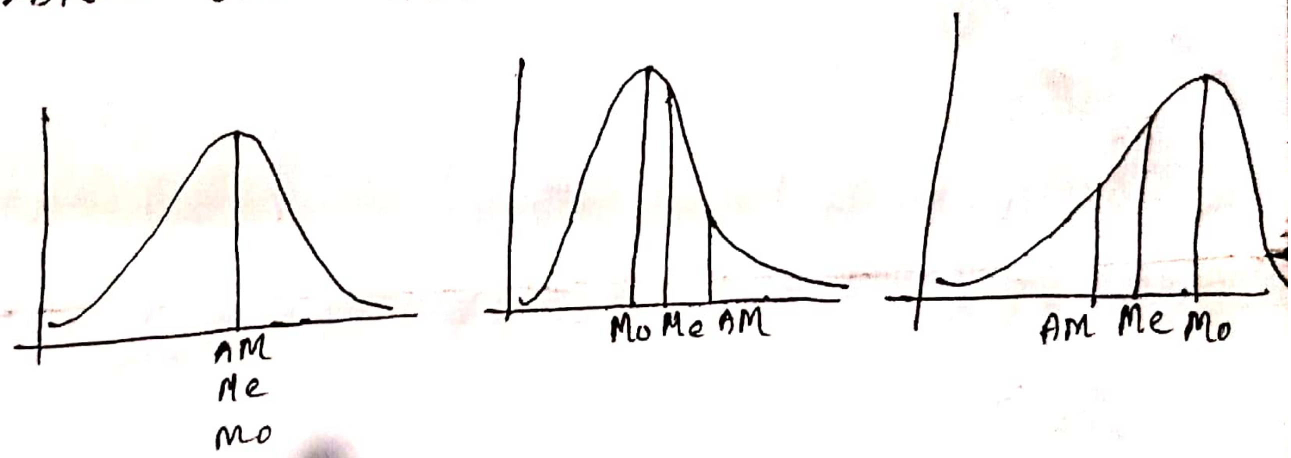
f_{Mo} : frequency of modal class

Relationship bet among Mean, Median and Mode :

When a distribution is symmetric, the Mo, Me and AM coincide.

When the AM exceeds the Mo, then the distribution is said to be positively skew or skew to the right.

When the AM is less than the Mo, then the distribution is said to be negatively skew or skew to the left.



The Geometric Mean :

The geometric mean G of n positive values $x_1, x_2, \ldots, x_n$ is defined as the nth positive root of the product of the values.

symbolically, $G = (x_1, x_2, \cdots x_n)^{1/n}$

The Harmonic Mean :

let $x_1, x_2, \cdots, x_n$ be a set of n positive values ($x_i > 0, \forall i$), then the HM of the values is given by :

$$HM = \frac{1}{\frac{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}{n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Scanned with CamScanner

i.e the HM is the reciprocal of the arithmatic mean of the reciprocals of the observations.

The HM is appropriate when we are dealing with rates, speeds, prices etc.