STATISTICS

MEAN DEVIATION

Mean Deviation

The mean deviation of a distribution is the average of the absolute deviations of the distribution's terms from its statistical mean (which can be the arithmetic mean, median, or mode).

- Mean deviation can be computed using any measure of central tendency. Nevertheless, mean deviation from the mean and median are frequently employed in statistical investigations.
- The mean deviation around the median is minimized.

1 Mean deviation for ungrouped data

Consider x_1, x_2, x_3, x_n as n values of a variable X, and let k be the statistical mean (Arithmetic Mean, Median, Mode) for which we need to determine the mean deviation. The mean deviation (M.D.) around k is calculated by:

$$\text{M.D. (k)} = \frac{|x_1 - k| + |x_2 - k| + |x_3 - k| + ... + |x_n - k|}{n} = \frac{\sum_{i=1}^{n} |x_i - k|}{n}$$

Ex.1 Determine the mean deviation around the mean for the given data:

12, 3, 18, 17, 4, 9, 17, 19, 20, 15, 8, 17, 2, 3, 16, 11, 3, 1, 0, 5

Sol. We have, $\bar{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{200}{20} = 10$

The absolute values of the deviations from the mean, specifically, $|x_i - \bar{x}|$ are

2, 7, 8, 7, 6, 1, 7, 9, 10, 5, 2, 7, 8, 7, 6, 1, 7, 9, 10, 5

Therefore $\sum_{i=1}^{20} |x_i - \bar{x}| = 124$

and $\text{M.D.}(\bar{x}) = \frac{124}{20} = 6.2$

Ex.2 The marks obtained by 9 students in an examination are as follows. Calculate their mean deviation from the median.

49, 68, 21, 32, 54, 38, 41, 66, 59

Sol. Organizing the observations in ascending order,

We have, 21, 32, 38, 41, 49, 54, 59, 66, 68

Number of observations = 9

Median = 5th term = 49

Calculation of mean deviation:

x_i	$\lvert d_i\rvert = \lvert x_i - 49\rvert$
21	28
32	17
38	11
41	8
49	0
54	5
59	10
66	17
68	19
Total	**115**

$$\text{M.D.} = \frac{1}{n}\sum |d_i| = \frac{115}{9} = 12.78$$

Ex.3 If $\bar{x}$ is the mean and the mean deviation from the mean is M. D($\bar{x}$), determine the number of observations lying between $\bar{x}$ – M.D. ($\bar{x}$) And $\bar{x}$ + M.D. ($\bar{x}$) For the following data:

34, 66, 30, 38, 44, 50, 40, 60, 42, 51.

Sol. The provided data, arranged in ascending order, is:

30, 34, 38, 40, 42, 44, 50, 51, 60, 66.

$$\bar{x} = \frac{30+34+38+40+42+44+50+51+60+66}{10}$$

$$\bar{x} = \frac{455}{10} = 45.5$$

Now, $|x_i - \bar{x}| = 15.5,\ 11.5, 7.5, 5.5,\ 3.5, 1.5, 4.5, 5.5, 14.5, 20.5$

Mean deviation from the mean

$$\text{M.D.}(\bar{x}) = \frac{\sum |(x_i - \bar{x})|}{10}$$

$$= \left(\frac{1}{10}\right)[15.5+11.5+7.5+5.5+3.5+1.5+4.5+5.5+ 14.5+20.5]$$

$$= \frac{1}{10}[90.0] = 9$$

Now, $\bar{x} - \text{M.D.}(\bar{x}) = 45.5 - 9 = 36.5$

And $\bar{x} + \text{M.D.}(\bar{x}) = 45.5 + 9 = 54.5$

Given observations which lie between 36.5 and 54.5 are 38, 40, 42, 44, 50, 51, which are six in number.

Six entries of given data lie between $\bar{x}$ – M.D. ($\bar{x}$) and $\bar{x}$ + M.D. ($\bar{x}$)

2 MEAN DEVIATION FOR GROUPED DATA

(a) Discrete Frequency Distribution:

Consider $x_1, x_2, x_3, \ldots, x_n$ be n observations with corresponding frequencies $f_1, f_2, f_3, \ldots, f_n$ and let K be the statistical mean (Arithmetic Mean, Median, Mode). The mean deviation (M.D.) around K is calculated by:

$$\text{M.D.}(k) = \frac{|x_1 - k|f_1 + |x_2 - k|f_2 + |x_3 - k|f_3 + \ldots + |x_n - k|f_n}{f_1 + f_2 + f_3 + \ldots + f_n}$$

$$= \frac{\sum_{i=1}^{n}|x_i - k|f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n}|d_i|f_i}{N},$$

Where $d_i = x_i - k$ and $N = \sum_{i=1}^{n} f_i$ = total frequency

The mean of given discrete frequency distribution is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_nx_n}{f_1 + f_2 + f_3 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

To determine the median of a given discrete frequency distribution, arrange the observations in ascending order. Subsequently, calculate the cumulative frequencies. Identify the observation whose cumulative frequency is equal to or just greater than $\frac{N}{2}$, where N is the total number of observations. This identified observation, situated in the middle of the data, represents the required median.

Ex.4 Calculate the mean deviation from the median for the given frequency distribution:

Age (in years)	10	12	15	18	21	23
Frequency	3	5	4	10	8	4

Sol. Calculating the mean deviation from the median.

Age(in years) x_i	Frequency (f_i)	Cumulative frequency	$f_i x_i$	$\|x_i - 18\|$	$F_i \|x_i - 18\|$
10	3	3	30	8	24
12	5	8	60	6	30
15	4	12	60	3	12
18	10	22	180	0	0
21	8	30	168	3	24
23	4	34	92	5	20
Total	N = 34				110

Here, $N = \sum f_i = 34, \frac{N}{2} = 17$ and $\frac{N}{2} + 1 = 18$

$$\text{Median} = \frac{(\text{value of } 17\text{th term}) + (\text{value of } 18\text{th term})}{2}$$

$$= \frac{18 + 18}{2} = 18 \text{ years}$$

$$\text{Mean deviation from the median} = \frac{\sum f_i |x_i - 18|}{N}$$

$$= \frac{110}{34} = 3.23 \text{ years}$$

Ex.5 The mean of 4, 7, 2, 8, 6, and \(a\) is 7. Determine the mean deviation about the median of these observations.

Sol. In this case, the count of observations, denoted as n, is 6.

Given $$\frac{4 + 7 + 2 + 8 + 6 + a}{6} = 7$$

or $$27 + a = 42$$

or $a = 15$

$\therefore$ Arranging the observations in ascending order, we get 2, 4, 6, 7, 8, 15

$\therefore$ Median, k = mean of $\frac{n}{2}^{th}$ observation

and $\left(\frac{n}{2}+1\right)^{th}$ observation $= \frac{3\text{rd observations} + 4\text{th observation}}{2}$

$$= \frac{6+7}{2} = 6.5$$

Calculation of mean:.

x_i	$\|x_i - k\|$
2	4.5
4	2.5
6	0.5
7	0.5
8	1.5
15	8.5
Total	**18**

Mean deviation about median $= \frac{\sum |x_i - k|}{n} = \frac{18}{6} = 3$

Ex.6 Determine the mean deviation around the mean for the provided data.

x_i	1	4	9	12	13	14	21	22
f_i	3	4	5	2	4	5	4	3

Sol. To calculate the mean and subsequently find the deviation around the mean, we create the following table:

x_i	f_i	$x_i f_i$	$\| x_i - \bar{x}\|$	$f_i\| x_i - \bar{x}\|$
1	3	3	11	33
4	4	16	8	32
9	5	45	3	15
12	2	24	0	0
13	4	52	1	4
14	5	70	2	10
21	4	84	9	36
22	3	66	10	30
Total	**30**	**360**		**160**

Mean $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{360}{30} = 12$

M.D. (about mean) $= \frac{\sum f_i \,|\, x_i - \bar{x} \,|}{\sum f_i} = \frac{160}{30} = 5.33$

(b) Continuous Frequency Distribution:

The mean of a continuous frequency distribution is computed under the assumption that the frequency in each class is cantered at its midpoint.

If x_i represents the mid-value of the i^{th} class, fi is the frequency of the i^{th} class, and k is the statistical mean Arithmetic Mean, Median, Mode, then the mean deviation (M.D.) around k is expressed as:

$$\text{M.D.}(k) = = \frac{\sum_{i=1}^{n} |x_i - k| f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} |d_i| f_i}{N},$$

Where $d_i = x_i - k$

And $N = \sum_{i=1}^{n} f_i$ = total frequency

Shortcut method for calculating mean deviation about mean

For the provided continuous frequency distribution, the arithmetic mean can be computed using the shortcut (step-deviation) method. The remaining steps of the procedure remain unchanged. In this approach,

1. Select a presumed mean (either the midpoint or a value close to it).
2. Compute the deviations of the observations (or midpoints of classes) from the chosen assumed mean.
3. If all the deviations share a common factor, divide each of them by this common factor to simplify the deviations.
4. The expression for the arithmetic mean when employing the step-deviation method is now formulated as:

$$\bar{x} = a + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times h,$$

Where $d_i = \frac{x_i - a}{h}$

a = assumed mean,

h = common factor

and $N = \sum_{i=1}^{n} f_i$

To calculate the median for a continuous frequency distribution

1. Compute $\frac{N}{2}, \left(N = \sum_{i=1}^{n} f_i\right)$.

2. The class that corresponds to a cumulative frequency just exceeding $\frac{N}{2}$ is identified as the median class.

3. $\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$,

Where, l = lower limit of median class

f = frequency of the median class

h = width of the median class

c = cumulative frequency of the class just preceding the median class

Ex.7 Calculate the mean deviation from the mean for the given dataset:

Classes	Frequency
0-100	4
100-200	8
200-300	9
300-400	10
400-500	7
500-600	5
600-700	4
700-800	3

Sol. We create the table as follows:

Classes	x_i	$d_i = \frac{x_i - 350}{100}$	f_i	$f_i d_i$	$\lvert x_i - k\rvert$	$f_i \lvert x_i - k\rvert$
0-100	50	−3	4	−12	308	1232
100-200	150	−2	8	−16	208	1664
200-300	250	−1	9	−9	108	972
300-400	350	0	10	0	8	80
400-500	450	1	7	7	92	644
500-600	550	2	5	10	192	960
600-700	650	3	4	12	292	1168
700-800	750	4	3	12	392	1176
Total			$N = \sum f_i = 50$	$\sum f_i d_i = 4$		7896

Actual mean $$k = A + \frac{\sum f_i d_i}{\sum f_i} \times h,$$

Where, A = assumed mean and h = class interval

$$k = 350 + \frac{4}{50} \times 100 = 358$$

Now, mean deviation,

$$\text{M.D.}(k) = \frac{\sum f_i \lvert x_i - k\rvert}{\sum f_i} = \frac{7896}{50} = 157.92$$

Ex.8 Calculate the mean deviation of the given distribution from the median.

Classes	10-20	20-30	30-40	40-50	50-60	60-70
Frequencies	10	12	8	16	14	10

Sol. We have

Classes	x_i	f_i	Cumulative frequency	$\lvert x_i - 43.125\rvert$	$f_i \lvert x_i - 43.125\rvert$
10-20	15	10	10	28.125	281.250
20-30	25	12	22	18.125	217.500
30-40	35	8	30	8.125	65.000
40-50	45	16	46	1.875	300.000
50-60	55	14	60	11.875	166.250
60-70	65	10	70	21.875	218.750
Total		N = 70			1248.750

Here, $$N = \sum f_i = 70, \ \frac{N}{2} = 35$$

The class with a cumulative frequency just exceeding 35 is 40-50, making it the median class.

Now, Median $$M = l + \frac{\frac{N}{2} - C}{f} \times h,$$

Here, l =40, N = 70, C = 30, h = 10, f = 16

$$M = 40 + \frac{35-30}{16} \times 10 = 43.125$$

Mean deviation from median $= \frac{\sum f_i \,|\, x_i - 43.125 \,|}{N}$

$$= \frac{1248.750}{70} = 17.83$$

1. The total of absolute deviations from the mean is higher than the total of absolute deviations from the median. In fact, the mean deviation about the median is the least. Therefore, the mean deviation about the mean is not very appropriate.
2. In a series characterized by a high degree of variability, the median does not serve as a representative measure of central tendency. Similarly, the mean deviation about the mean is not a highly reliable measure of dispersion in such cases.
3. Mean deviation is computed using the absolute values of deviations, making it unsuitable for further algebraic manipulation.

Important formula / points

- The smallest mean deviation occurs around the median.
- Mean deviation

Ungrouped data: $\text{M.D.} = \frac{\sum_{i=1}^{n} |x_i - k|}{n}$; K be the statistical mean (A.M., median, mode)

Discrete / continuous frequency distribution: $\text{M.D.} = \frac{\sum_{i=1}^{n} |x_i - k| f_i}{\sum_{i=1}^{n} f_i}$

Where, x_i = observations / mid – value of i^{th} class

f_i = frequency of i^{th} observation / class

n = number of observations

k = statistical mean (A.M., Median, Mode)