MEASURES OF CENTRAL TENDENCY AND DISPERSION

(KEY CONCEPTS + SOLVED EXAMPLES)

MEASURES OF CENTRAL TENDENCY AND DISPERSION

1.Measures of Central Tendency

2.Arithmetic Mean

3.Geometric Mean

4.Harmonic Mean

5.Median

6.Mode

7.Measures of dispersion

8.Variance and standard deviation

9.Mean and variance of binomial distribution

KEY CONCEPTS

1. Measures of Central Tendency

It is that single value which may be taken as the most suitable representative of the data. This single value is known as the average. Averages are generally, the central part of the distribution and therefore, they are also called the measures of **Central Tendency.**

It can be divided into two groups :

(a) MATHEMATICAL AVERAGE :

i. Arithmetic mean or mean

ii. Geometric mean

iii. Harmonic mean

(b) POSITIONAL AVERAGE :

i. Median

ii. Mode or positional average

2. Arithmetic Mean

2.1 Individual observation or unclassified data :

If $x_1, x_2, \ldots\ldots x_n$ be n observations, then their arithmetic mean is given by

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

2.2 Arithmetic mean of discrete frequency distribution :

Let $x_1, x_2, \ldots\ldots, x_n$ be n observation and let $f_1, f_2, \ldots\ldots, f_n$ be their corresponding frequencies, then their mean

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

2.2.1

Short cut method : If the values of x or (and) f are large the calculation of arithmetic mean by the previous formula used, is quite tedious and time consuming. In such case we take the deviation from an arbitrary point A.

$$\bar{x} = A + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

where A = Assumed mean

$d_i = x_i - A$ = deviation for each term

2.2.2

Step deviation method : Sometimes during the application of shortcut method of finding the mean, the deviation d_i are divisible by a common number h (say). In such case the arithmetic is reduced to a great extent taken by

$$u_i = \frac{x_i - A}{h}, \; i = 1, 2, \ldots\ldots n$$

$$\therefore \text{ mean } \bar{x} = A + h\left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right)$$

2.3 Weighted arithmetic mean :

If $w_1, w_2, w_3, \ldots\ldots, w_n$ are the weight assigned to the values $x_1, x_2, x_3, \ldots\ldots, x_n$ respectively, then the weighted average is defined as -

Weighted A.M. = $\dfrac{w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n}{w_1 + w_2 + \ldots\ldots + w_n}$ $\quad \overline{x} = \dfrac{\sum_{i=1}^{n} w_ix_i}{\sum_{i=1}^{n} w_i}$

2.4 Combined mean :

If $\overline{x}_1, \overline{x}_2, \ldots\ldots, \overline{x}_k$ are the mean of k series of sizes $n_1, n_2, \ldots, n_k$ respectively then the mean of the composite series is given by

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2 + \ldots\ldots n_k\overline{x}_k}{n_1 + n_2 + \ldots\ldots + n_k}$$

2.5 Properties of Arithmetic Mean :

1. In a statistical data, the sum of the deviation of items from A.M. is always zero.

i.e. $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$

or $\sum_{i=1}^{n} x_i - n\overline{x}$

or $n\,\overline{x} - n\,\overline{x} = 0$

$\left(\because \overline{x} = \dfrac{\sum x_i}{n} \right)$

2. In a statistical data, the sum of squares of the deviation of items from A.M. is least i.e. $\sum_{i=1}^{n}(x_i - \overline{x})^2$ is least

3. If each of the n given observation be doubled, then their mean is doubled.

4. If $\overline{x}$ is the mean of $x_1, x_2, \ldots\ldots, x_n$. The mean of $ax_1, ax_2, \ldots\ldots, ax_n$ is a $\overline{x}$ where a is any number different from zero.

5. Arithmetic mean is independent of origin i.e. it is not effected by any change in origin.

2.6 MERITS AND DEMERITS OF ARITHMETIC MEAN :

2.6.1 Merits of Arithmetic Mean :

i. It is rigidly defined.

ii. It is based on all the observation taken.

iii. It is calculated with reasonable ease and rapidity.

iv. It is least affected by fluctuations in sampling.

v. It is based on each observation and so it is a better representative of the data.

vi. It is relatively reliable.

vii. Mathematical analysis of mean is possible.

2.6.2 Demerits of Arithmetic Mean :

i. It is severely affected by the extreme values.

ii. It cannot be represented in the actual data since the mean does not coincide with any of the observed value.

iii. It cannot be computed unless all the items are known.

iv. It cannot be calculated for qualitative data incapable of numerical measurements.

v. It cannot be used in the study of ratios, rates etc.

3. Geometric Mean

3.1 Individual data : If x_1, x_2, x_3,......, x_n are n values of a variate x, none of them being zero, then the geometric mean G is defined as-

$$G = (x_1\, x_2\, x_3 \,.......x_n)^{1/n}$$

or $$G = \text{antilog}\left(\frac{\log x_1 + \log x_2 + ...+ \log x_n}{n}\right)$$

or $$G = \text{antilog}\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)$$

3.2 Geometric Mean of grouped data :

Let x_1, x_2, ...,x_n be n observation and let f_1, f_2 ,......,f_n be their corresponding frequency then their Geometric Mean is

$$G = \left(x_1^{f_1} x_2^{f_2} ....x_n^{fn}\right)^{1/N}$$

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

$$\therefore\ G = \text{antilog}\left(\frac{\sum_{i=1}^{n} f_i \log x_i}{\sum_{i=1}^{n} f_i}\right)$$

4. Harmonic Mean

Harmonic mean is reciprocal of mean of reciprocal.

4.1 Individual observation :

The H.M. of x_1, x_2,........,x_n of n observation is given by

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ...+ \frac{1}{x_n}}$$

i.e. $$H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

4.2 H.M. of grouped data :

Let x_1, x_2, ..., x_n be n observation and let f_1, f_2,.........,f_n be their corresponding frequency then H.M. is

$$H = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)}$$

Relation between A.M. , G.M., and H.M.

A.M. ≥ G.M.≥ H.M.

Equality sign holds only when all the observations in the series are same.

5. Medain

Median is the middle most or the central value of the variate in a set of observations, when the observations are arranged either in ascending or in descending order of their magnitudes. It divides the arranged series in two equal parts.

5.1 Median of an individual series :

Let n be the number of observations-

(i) arrange the data in ascending or descending order.

(ii) **(a) if n is odd then-**

$$\text{Median (M)} = \text{Value of } \left(\frac{n+1}{2}\right)^{th} \text{ observation}$$

(b) If n is even then

$$\text{Median (M)} = \text{mean of } \left(\frac{n}{2}\right)^{th} \text{ and } \left(\frac{n}{2}+1\right)^{th} \text{ observation}$$

$$\text{i.e. } M = \frac{\left(\frac{n}{2}\right)^{th}\text{observation} + \left(\frac{n}{2}+1\right)^{th}\text{observation}}{2}$$

5.2 Median of the discrete frequency distribution:

Algorithm to find the median :

Step - I : Find the cumulative frequency (C. F.)

Step- II : Find $\frac{N}{2}$, where $N = \sum_{i=1}^{n} f_i$

Step -III : See the cumulative frequency (C.F.) just greater than $\frac{N}{2}$ and determine the corresponding value of the variable.

Step-IV : The value obtained in step III is the median.

5.3 Median of grouped data or continuous series:

Let the no. of observations be n

(i) Prepare the cumulative frequency table

(ii) Find the median class i.e. the class in which the $\left(\frac{N}{2}\right)^{th}$ observation lies

(iii) The median value is given by the formulae

$$\text{Median (M)} = \ell + \left[\frac{\left(\frac{N}{2}\right) - F}{f}\right] \times h$$

N = total frequency = Σf_i

ℓ = lower limit of median class

f = frequency of the median class

F = cumulative frequency of the class preceding the median class

h = class interval (width) of the median class

5.4 Properties of Median :

(A) The sum of the absolute value of deviations of the items from median is minimum

(B) It is a positional average and it is not influenced by the position of the items.

6. Mode

Mode is that value in a series which occurs most frequently. In a frequency distribution, mode is that variate which has the maximum frequency.

6.1 Computation of Mode :

(i) Mode for individual series : In the case of individual series, the value which is repeated maximum number of times is the mode of the series.

6.2 Mode for grouped data (discrete frequency distribution series)

In the case of discrete frequency distribution, mode is the value of the variate corresponding to the maximum frequency.

6.3 Mode for continuous frequency distribution :

(i) First find the model class i.e. the class which has maximum frequency. The model class can be determined either by inspecting or with the help of grouping data.

(ii) The mode is given by the formula

$$\text{Mode} = \ell + \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}} \times h$$

where $\ell \rightarrow$ lower limit of the model class

$h \rightarrow$ width of the model class

$f_m \rightarrow$ frequency of the model class

$f_{m-1} \rightarrow$ frequency of the class preceding model class

$f_{m+1} \rightarrow$ frequency of the class succeeding model class

(iii) In case the model value lies in a class other than the one containing maximum frequency (model class) then we use the following formula :-

$$\text{Mode} : \ell + \frac{f_{m+1}}{f_{m-1} + f_{m+1}} \times h$$

6.4 Properties of Mode : It is not effected by presence of extremely large or small items.

Relationship between mean, mode and median :

(i) In symmetrical distribution

Mean = Mode = Median

(ii) In skew (moderately symmetrical) distribution

Mode = 3 median – 2 mean

7. Measures of Dispersion

Dispersion is the measure of the variations. The degree to which numerical data tend to spread about an average value is called the dispersion of the data.

The measures of dispersion commonly used are:

(i) Range

(ii) Quartile deviation or the semi- interquartile range

(iii) Mean Deviation

(iv) Standard Deviation

Here we will discuss the mean deviation and standard deviation.

Mean Deviation : Mean deviation is defined as the arithmetic mean of the absolute deviations of all the values taken about any central value.

i. Mean deviation of individual observations:

If $x_1, x_2, ..., x_n$ are n values of a variable x, then the mean deviation from an average A (median or AM) is given by

$$\text{M.D.} = \frac{1}{n}\sum_{i=1}^{n}|x_i - A|$$

$$= \frac{1}{n}\sum|d_i|, \text{ where } d_i = x_i - A$$

ii. Mean deviation of a discrete frequency distribution :

If $x_1, x_2,....., x_n$ are n observation with frequencies $f_1, f_2, ..., f_n$, then mean deviation from an average A is given by -

$$\text{Mean Deviation} = \frac{1}{N}\sum f_i|x_i - A|$$

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

iii. Mean deviation of a grouped or continuous frequency distribution:

For calculating mean deviation of a continuous frequency distribution the procedure is same as for a discrete frequency distribution. The only difference is that here we have to obtain the mid-point of the various classes and take the deviations of these mid points from the given central value (median or mean)

8. Variance and Standard Deviation

The variance of a variate x is the arithmetic mean of the squares of all deviations of x from the arithmetic mean of the observations and is denoted by var (x) or σ^2

The positive square root of the variance of a variate x is known as standard deviation i.e. standard deviation = + $\sqrt{\text{var}(x)}$

(i) Variance of Individual observations :

If $x_1, x_2,....., x_n$ are n values of a variable x, then by definition

$$\text{var}(x) = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \sigma^2 \qquad ...(i)$$

$$\text{or var}(x) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 \qquad ...(ii)$$

If the values of variable x are large, the calculation of variance from the above formulae is quite tedious and time consuming. In that case, we take deviation from an arbitrary point A (say) then

$$\text{var}(x) = \frac{1}{n}\sum_{i=1}^{n}d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}d_i\right)^2 \qquad ...(iii)$$

(ii) Variance of a discrete frequency distribution:

If $x_1, x_2,......, x_n$ are n observations with frequencies $f_1, f_2,.....,f_n$ then $\text{var}(x) = \frac{1}{N}\left\{\sum_{i=1}^{n}f_i(x_i - \bar{x})^2\right\}$ ...(i) or var (x) = $\frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \bar{x}^2$ ...(ii)

If the values of x or f are large, we take the deviations of the values of variable x from an arbitrary point A. (say)

$\therefore$ $d_i = x_i - A$; i = 1, 2, ...n

$$\therefore \text{ Var }(x) = \frac{1}{N}\left(\sum_{i=1}^{n} f_i\, d_i^{\,2}\right) - \left(\frac{1}{N}\sum_{i=1}^{n} f_i\, d_i\right)^2 \qquad ...(iii)$$

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

Sometime $d_i = x_i - A$ are divisible by a common number h (say)

then

$u_i = \frac{x_i - A}{h} = \frac{d_i}{h}$, i = 1, 2,....., n

then

$$\text{var }(x) = h^2\left[\frac{1}{N}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{N}\sum_{i=1}^{n} f_i u_i\right)^2\right] \qquad ...(iv)$$

(iii) Variance of a grouped or continuous frequency distribution :

In a grouped or continuous frequency distribution any of the formulae discussed in discrete frequency distribution can be used.

9. Mean and of Variance of Binomial Distribution

If the frequencies of the values 0, 1, 2,...., n of a variate are represented by the following coefficients of a binomial :

q^n, ${}^nC_1\, q^{n-1}p$, ${}^nC_2 q^{n-2}\, p^2$,...., p^n

where p is the probability of the success of the experiment (variate), q is the probability of its failure and p + q = 1 i.e. distribution is a binomial distribution : then

$P(x = r) = {}^nC_r\, q^{n-r}\, p^r$

mean $\bar{x} = \sum p_i x_i = np$

Variance $\sigma^2 = npq = \bar{x}\, q$

SOLVED EXAMPLES

Ex.1 Mean of 25 observations was found to be 78.4. But later on it was found that 96 was misread 69. The correct mean is

(A) 79.24 (B) 79.48 (C) 80.10 (D) None of these

Sol. Mean $\overline{x} = \frac{\Sigma x}{n}$ or $\Sigma x = n\,\overline{x}$

$\Sigma x = 25 \times 78.4 = 1960$

But this Σx is incorrect as 96 was misread as 69.

$\therefore$ correct $\Sigma x = 1960 + (96 - 69) = 1987$ $\therefore$ correct mean $= \frac{1987}{25} = 79.48$ **Ans. [B]**

Ex.2 If $\overline{x}$ is the mean of $x_1, x_2, \ldots, x_n$ then mean of $x_1 + a, x_2 + a, \ldots, x_n + a$ where a is any number positive or negative is-

(A) $\overline{x} + a$ (B) $\overline{x}$ (C) $a\,\overline{x}$ (D) None of these

Sol. We have $\overline{x}' = \frac{x_1 + x_2 + \ldots + x_n}{n}$

Let $\overline{x}$ be the mean of $x_1 + a, x_2 + a, \ldots, x_n + a$ then

$$\overline{x}' = \frac{(x_1 + a) + (x_2 + a) + \ldots + (x_n + c)}{n} = \frac{(x_1 + x_2 + \ldots + x_n) \ldots + na}{n}$$

$$= \frac{x_1 + x_2 + \ldots + x_n}{n} + a = \overline{x} + a$$

Ans. [A]

Ex.3 Mean wage from the following data

Wage (In Rs.)	*800*	*820*	*860*	*900*	*920*	*980*	*1000*
No. of workers	7	14	19	25	20	10	5

is

(A) Rs.889 (B) Rs. 890.4 (C) Rs.891.2 (D) None of these

Sol. Let the assumed mean be, A = 900.The given data can be written as under :

Wage (in Rs.) x_i	**No. of workers** f_i	$d_i = x_i - A$ $= x_i - 900$	$u_i = \frac{x_i - 900}{20}$	$f_i u_i$
800	7	– 100	– 5	– 35
820	14	– 80	– 4	– 56
860	19	– 40	– 2	– 38
900	25	0	0	0
920	20	20	1	20
980	10	80	4	40
1000	5	100	5	25
	N $= \Sigma f_1 = 100$		$\Sigma f_1 u_i = -44$	

Here A = 900, h = 20

$\therefore$ Mean = $\overline{X} = A + h\left(\frac{1}{N}\sum f_i\ u_i\right) = 900 + 20\left(-\frac{44}{100}\right) = 891.2$

Hence, mean wage = Rs. 891.2.

Ans. [C]

Ex.4 Median from the following distribution

Class	*5-10*	*10-15*	*15-20*	*20-25*	*25-30*	*30-35*	*35-40*	*40-45*
frequency	5	6	15	10	5	4	2	2

is-

(A) 19.0 (B) 19.2 (C) 19.3 (D) 19.5

Sol.

Class	**Frequency**	**Cumulative frequency**
5 - 10	5	5
10 - 15	6	11
15 - 20	15	26
20 - 25	10	36
25 - 30	5	41
30 - 35	4	45
35 - 40	2	47
40 - 45	2	49
	N = 49	

We have N = 49. $\therefore\ \frac{N}{2} = \frac{49}{2} = 24.5$

The cumulative frequency just greater than N/2, is 26 and the corresponding class is 15-20. Thus 15-20 is the median class such that

$\ell = 15,\ f = 15,\ F = 11,\ h = 5.$

$\therefore$ Median = $\ell + \frac{N/2-F}{f} \times h = 15 + \frac{24.5-11}{15} \times 5 = 15 + \frac{13.5}{3} = 19.5$ **Ans. [D]**

Ex.5 Mean deviation about mean from the following data :

x_i :	*3*	*9*	*17*	*23*	*27*
f_i :	8	10	12	9	5

is -

(A) 7.15 (B) 7.09 (C) 8.05 (D) None of these

Sol. Calculation of mean deviation about mean.

x_i	f_i	$f_i x_i$	$\lvert x_i - 15\rvert$	$f_i\ \lvert x_i - 15\rvert$
3	8	24	12	96
9	10	90	6	60
17	12	204	2	24
23	9	207	8	72

27	5	135	12	60
	N =Σ f_1 = 44	Σf_1x_1= **660**		$\Sigma f_1\lvert x_1 - 15\rvert$= **312**

Mean = $\overline{X} = \frac{1}{N}(\Sigma f_1x_1) = \frac{660}{44} = 15$

Mean deviation = M.D. = $\frac{1}{N}\Sigma f_1 \lvert x_1 - 15\rvert = \frac{312}{44} = 7.09$. **Ans.**

[B]

Ex.6 Variance of the data given below

size of item	*3.5*	*4.5*	*5.5*	*6.5*	*7.5*	*8.5*	*9.5*
frequency	3	7	22	60	85	32	8

is-

(A) 1.29 (B) 2.19 (C) 1.32 (D) None of these

Sol. Let the assumed mean be A = 6.5

Calculation of variance

size of item x_i	f_i	$d_i = x_i - 6.5$	d_i^2	f_id_i	$f_id_i^2$
3.5	3	–3	9	– 9	27
4.5	7	–2	4	–14	28
5.5	22	–1	1	–22	22
6.5	60	0	0	0	0
7.5	85	1	1	85	85
8.5	32	2	4	64	128
9.5	8	3	9	24	72
	N = Σf_i = 217			**Σf_id_i=128**	**$\Sigma f_id_i^2$ = 362**

Here, N =217, Σf_id_i =128 and $\Sigma f_id_i^2$ =362

$$\therefore \text{Var}(X) = \left(\frac{1}{N}\sum f_i d_i^2\right) - \left(\frac{1}{N}\sum f_i d_i\right)^2 = \frac{362}{217} - \left(\frac{128}{217}\right)^2 = 1.668 - 0.347 = 1.321$$

Ans. [C]