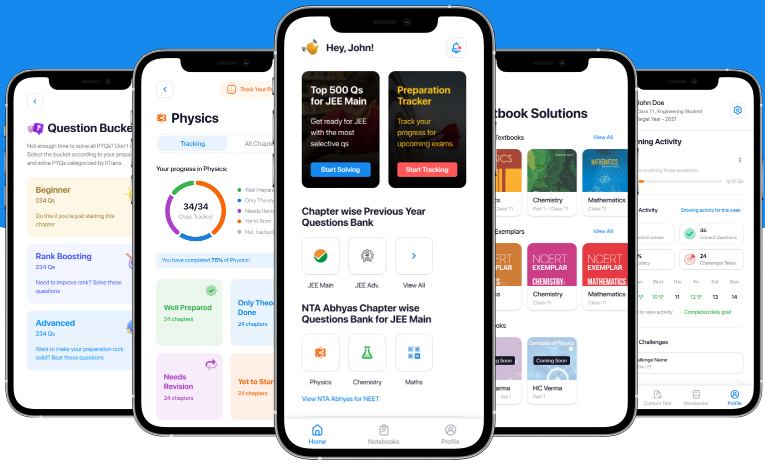
MARKS



The only app you need to prepare for

JEE MAIN | JEE ADV. | BITSAT | WBJEE | MHT CET | and more...

GET IT ON Google Play

4.8 Rating on Google Play

50,000+ Students using daily

2,00,000+ Questions available

With MARKS app you can do all these things for free 😍

- Solve Chapter-wise PYQ of JEE Main, JEE Advanced, NEET, BITSAT, WBJEE, MHT CET & more
- Create Unlimited Custom Tests for any exam
- Attempt Top Questions for JEE Main which can boost your rank
- Track your exam preparation with Preparation Trackers
- Complete daily goals, rank up on the leaderboard & compete with other aspirants

MARKS

4.8 Rating on Google Play

50,000+ Students using daily

2,00,000+ Questions available

GET IT ON Google Play

Statistics

MEASURES OF CENTRAL TENDENCY, DISPERSION

One of the most important objectives of statistical analysis is to get one single value that describes the characteristic of the entire data. Such a value is called the central value or an average.

The following are the important types of averages:

1. Arithmetic Mean
2. Geometric Mean
3. Harmonic mean
4. Median
5. Mode

We consider these measures in three cases (i) Individual series (i.e. each individual observation is given) (ii) discrete series (i.e the observations along with number of times a particular observation called the frequency is given) (iii) continuous series (i.e. the class intervals along with their frequencies are given)

ARITHMETIC MEAN

(i) Individual Series : If $x_1, x_2 \ldots\ldots x_n$ are the values of the variable x, then the arithmetic mean usually denoted by $\bar{x}$ is given by

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

(ii) Discrete Series : If a variable takes values $x_1, x_2 \ldots\ldots x_n$ with corresponding frequencies $f_1, f_2 \ldots\ldots f_n$ then the arithmetic mean $\bar{x}$ is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 \ldots\ldots + f_nx_n}{f_1 + f_2 \ldots + f_n} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i, \quad \text{where} \quad N = \sum_{i=1}^{n} f_i$$

(iii) Continuous Series : In case of a set of data with class intervals, we cannot find the exact value of the mean because we do not know the exact values of the variables. We, therefore, try to obtain an approximate value of the mean. The method of approximate is to replace all the observed values belonging to a class by mid-value of the class. If $x_1, x_2 \ldots x_n$ are the mid values of the class intervals having corresponding frequencies $f_1, f_2 \ldots f_n$ then we apply the same formula as in discrete series.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i, \quad N = \sum_{i=1}^{n} f_i$$

PROPERTIES OF ARITHMETIC MEAN

(i) Sum of all the deviations from arithmetic mean is zero i.e.,

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = 0 \text{ (in case of individual series)}$$

$$\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right) = 0 \text{ (in case of discrete or continuous series)}$$

(ii) If each observation is increased or decreased by a given constant K, the mean is also increased or decreased by K

The property is also known as effect of change of origin. K can be taken to be any number. However, to simplify the calculations, K should be taken as a value which is in the middle of the table.

(iii) Step Deviation Method or change of scale

If $x_1, x_2 ....x_n$ are mid values of class intervals with corresponding frequencies $f_1, f_2, ..., f_n$ then we may change the scale by taking $d_i = \frac{x_i - A}{h}$, in this case.

$$\bar{x} = A + h \times \left(\frac{1}{N} \sum f_i d_i \right)$$

A and h can be any numbers but if the lengths of class intervals are equal then h may be taken as width of the class interval.

In particular if each observation is multiplied or divided by a constant, the mean is also multiplied or divided by the same constant.

(iv) The sum of the squared deviation of the variate from their mean is minimum i.e., the quantity $\sum (x_i - A)^2$ or $\sum f_i (x_i - A)^2$ is minimum when $A = \bar{x}$

(v) If $\bar{x}_1$ and $\bar{x}_2$ be the means of two related groups having n_1 and n_2 items respectively then the combined mean $\bar{x}$ of both the groups is given by $\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$

WEIGHTED ARITHMETIC MEAN

If $x_1, x_2, ..., x_n$ are n vlaues of a variable X and $w_1, w_2, ..., w_n$ denote respectively their weights, then their weighted mean $\bar{X}_w$ is given by

$$\bar{X}_w = \frac{w_1 x_1 + w_2 x_2 + ... + w_n x_n}{w_1 + w_2 + ... + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

GEOMETRIC MEAN

In case of individual series $x_1, x_2 ....x_n$

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

In case of discrete or continuous series

$$\text{G.M.} = \left(x_1^{f_1} x_2^{f_2} .... x_n^{f_n} \right)^{1/N}, \text{ where } N = \sum_{i=1}^{n} f_i$$

HARMONIC MEAN

The harmonic mean is based on the reciprocals of the value of the variable

$$\text{H.M.} = \frac{1}{\frac{1}{n}\left(\frac{1}{x_i} + \frac{1}{x_2} + ... + \frac{1}{x_n} \right)} \quad \text{or} \quad \frac{1}{H} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \text{ (Incase of Individual series)}$$

and $\dfrac{1}{H} = \dfrac{1}{\dfrac{1}{N}\sum_{i=1}^{n} f_i \dfrac{1}{x_i}}$ (in case of discrete series or continuous series)

If $x_1, x_2, \ldots x_n > 0$ then it is known that A.M. $\geq$ G.M. $\geq$ H.M.

MEDIAN

It refers to the middle value in a distribution.

In case of individual series, in order to find median, arrange the data in ascending or descending order of magnitude. In case of odd number of values

Median = size of $\dfrac{n+1}{2}$th item. In case of even number of values Median = average of $\dfrac{n}{2}$th and $\dfrac{n+2}{2}$th observation.

In case of discrete frequency distribution the median is obtained by considering the cumulative frequency (c.f.). Find $\dfrac{N}{2}$, where $N = \sum_{i=1}^{n} f_i$. Find the cumulative frequency (c.f.) just more than N/2. The corresponding value of x is median.

In case of continuous distribution, the class corresponding to c.f. just more than N/2 is called the median class and the median is obtained by Median = $l + \dfrac{h}{f}\left(\dfrac{N}{2} - C\right)$

Where $l =$ the lower limit of the median class; $f =$ the frequency of the median class; $h =$ the width of the median class; $C =$ the c.f. of the class preceding to the median class and $N = \sum_{i=1}^{n} f_i$

MODE

The mode is that value in a series of observations which occurs with greatest frequency. In case of individual series, for determining the mode, count the number of times the various values repeat themselves and the value occurring the maximum number of times in the model value.

In case of discrete series, quite often mode can be determined just by inspection i.e. by looking to that value of variable around which the items are most heavily concentrated.

In case of continuous series,

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_2 - f_0} \times h$$

Where $l =$ the lower limit of the modal class i.e. the class having maximum frequency; 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 and h = width of the modal class.

When mode is ill-defined i.e. there are two or more values which occur with equal maximum frequency, the mode can be computed by

Mode = 3 median - 2 mean

DISPERSION

Literally, dispersion means 'scatteredness'. Dispersion measures the degree of scatteredness of the variable about a central value. Different measures of dispersion are

1. Range
2. Mean-deviation
3. Quartile deviation
4. Standard deviation

1. Range

The range is the difference between the largest and smallest observation.

2. Mean-deviation

If $x_1, x_2 \ldots\ldots\ldots\ldots x_n$ are n observations then mean deviation about a point A is given by

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

In case of discrete or continuous series

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

M.D. is least when taken from the median

Standard Deviation

Variance σ^2 in case of individual series is given by

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

If x_1, x_2, ……, x_n occur with frequency f_1, f_2….f_n respectively then

$$\sigma^2 \text{ (variance)} = \frac{1}{N}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{n}\sum f_i x_i\right)^2$$

and standard deviation =+ $\sqrt{\text{variance}}$

There is no effect of change of origin on standard deviation

$$\sigma_x^2 = h^2\left[\frac{1}{N}\sum f_i d_i^2 - \left(\frac{1}{N}\sum f_i d_i\right)^2\right]$$

If there are two samples of sizes n_1 and n_2 with $\overline{x_1}$ and $\overline{x_2}$ as their means σ_1 and σ_2 their standard deviations respectively, then the combined variance is given by

$$\sigma^2 = \frac{n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)}{n_1 + n_2}$$

where $d_1 = \overline{x_1} - \bar{x}$ and $d_2 = \overline{x_2} - \bar{x}$, $\bar{x}$ being the combined mean.

COEFFICIENT OF VARIATION

$$= \frac{\sigma}{X} \times 100$$

Properties of Variance :

1. Let $x_1, x_2, x_3, \ldots\ldots, x_n$ be n values of a variable X. If these values are changed to $x_1 + a$, $x_2 + a$, ..., $x_n + a$, where $a \in R$, then the variance remains unchanged.
2. Let x_1, x_2, ..., x_n values of a variable X and let 'a' be a non-zero real number. Then, the variance of the observations ax_1, ax_2, ..., ax_n is a^2 Var (X).
3. Let x_1, x_2, x_3, ...,x_n be n values of a variable X, and let $x_i = a + hu_i$, i = 1,2, ...,n, where u_1, u_2,.....,u_n are the values of variable U. Then, Var (X) = h^2 Var (U), $h \neq 0$