Testprep Kart

Knowledge... Everywhere

Mathematics

Measure of Central Tendency

Table of Content

1. Introduction.

An average or a central value of a statistical series in the value of the variable which describes the characteristics of the entire distribution.

The following are the five measures of central tendency.

(1) Arithmetic mean
(2) Geometric mean
(3) Harmonic mean
(4) Median
(5) Mode

2. Arithmetic Mean.

Arithmetic mean is the most important among the mathematical mean.
According to Horace Secrist,
"The arithmetic mean is the amount secured by dividing the sum of values of the items in a series by their number."

(1) **Simple arithmetic mean in individual series (Ungrouped data)**

(i) **Direct method:** If the series in this case be $x_1, x_2, x_3, ..., x_n$ then the arithmetic mean $\bar{x}$ is given by

$$\bar{x} = \frac{\text{Sum of the series}}{\text{Number of terms}}, \text{ i.e., } \bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(ii) **Short cut method**

Arithmetic mean $(\bar{x}) = A + \frac{\Sigma d}{n}$,

Where, A = assumed mean, d = deviation from assumed mean = x – A, where x is the individual item, Σd = sum of deviations and n = number of items.

(2) **Simple arithmetic mean in continuous series (Grouped data)**

(i) **Direct method:** If the terms of the given series be $x_1, x_2, ..., x_n$ and the corresponding frequencies be $f_1, f_2, ... f_n$, then the arithmetic mean $\bar{x}$ is given by, $\bar{x} = \frac{f_1 x_1 + f_2 x_2 + ... + f_n x_n}{f_1 + f_2 + ... + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$.

(ii) **Short cut method:** Arithmetic mean $(\bar{x}) = A + \frac{\sum f(x - A)}{\sum f}$

Where A = assumed mean, f = frequency and x – A = deviation of each item from the assumed mean.

(3) **Properties of arithmetic mean**

(i) Algebraic sum of the deviations of a set of values from their arithmetic mean is zero. If x_i / f_i, i = 1, 2, ..., n is the frequency distribution, then

$\sum_{i=1}^{n} f_i (x_i - \bar{x}) = 0$, $\bar{x}$ being the mean of the distribution.

(ii) The sum of the squares of the deviations of a set of values is minimum when taken about mean.

(iii) **Mean of the composite series :** If $\bar{x}_i, (i = 1, 2, ..., k)$ are the means of k-component series of sizes $n_i, (i = 1, 2, ..., k)$ respectively, then the mean $\bar{x}$ of the composite series obtained on combining the component series is given by the formula $\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + ... + n_k \bar{x}_k}{n_1 + n_2 + ... + n_k} = \sum_{i=1}^{n} n_i \bar{x}_i \Big/ \sum_{i=1}^{n} n_i$.

3. Geometric Mean.

If $x_1, x_2, x_3, ..., x_n$ are n values of a variate x, none of them being zero, then geometric mean (G.M.) is given by $\text{G.M.} = (x_1.x_2.x_3......x_n)^{1/n} \Rightarrow \log(\text{G.M.}) = \frac{1}{n}(\log x_1 + \log x_2 + ... + \log x_n)$.

In case of frequency distribution, G.M. of n values $x_1, x_2, ...x_n$ of a variate x occurring with frequency $f_1, f_2, ..., f_n$ is given by $\text{G.M.} = (x_1^{f_1}.x_2^{f_2}.....x_n^{f_n})^{1/N}$, where $N = f_1 + f_2 + ... + f_n$.

4. Harmonic Mean.

The harmonic mean of n items $x_1, x_2, ..., x_n$ is defined as $\text{H.M.} = \dfrac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + ... + \dfrac{1}{x_n}}$.

If the frequency distribution is $f_1, f_2, f_3, ..., f_n$ respectively, then $\text{H.M.} = \dfrac{f_1 + f_2 + f_3 + ... + f_n}{\left(\dfrac{f_1}{x_1} + \dfrac{f_2}{x_2} + ... + \dfrac{f_n}{x_n}\right)}$

Note: A.M. gives more weightage to larger values whereas G.M. and H.M. give more weightage to smaller values.

5. Median.

Median is defined as the value of an item or observation above or below which lies on an equal number of observations i.e., the median is the central value of the set of observations provided all the observations are arranged in the ascending or descending orders.

(1) **Calculation of median**

(i) **Individual series:** If the data is raw, arrange in ascending or descending order. Let n be the number of observations.

If n is odd, Median = value of $\left(\frac{n+1}{2}\right)^{th}$ item.

If n is even, Median = $\frac{1}{2}\left[\text{value of}\left(\frac{n}{2}\right)^{\text{th}}\text{ item} + \text{value of}\left(\frac{n}{2}+1\right)^{\text{th}}\text{ item}\right]$

(ii) **Discrete series:** In this case, we first find the cumulative frequencies of the variables arranged in ascending or descending order and the median is given by

Median = $\left(\frac{n+1}{2}\right)^{th}$ observation, where n is the cumulative frequency.

(iii) **For grouped or continuous distributions:** In this case, following formula can be used

(a) For series in ascending order, Median = $l + \frac{\left(\frac{N}{2}-C\right)}{f} \times i$

Where l = Lower limit of the median class

f = Frequency of the median class

N = The sum of all frequencies

I = The width of the median class

C = The cumulative frequency of the class preceding to median class.

(b) For series in descending order

Median = $u - \left(\frac{\frac{N}{2}-C}{f}\right) \times i$, where u = upper limit of the median class

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

As median divides a distribution into two equal parts, similarly the quartiles, quantiles, deciles and percentiles divide the distribution respectively into 4, 5, 10 and 100 equal parts. The j^{th} quartile is given by $Q_j = l + \left(\frac{j\frac{N}{4}-C}{f}\right) i; j = 1, 2, 3$. Q_1 is the lower quartile, Q_2 is the median and Q_3 is called the upper quartile.

(2) **Lower quartile**

(i) **Discrete series:** $Q_1 = \text{size of} \left(\frac{n+1}{4}\right)^{\text{th}} \text{ item}$

(ii) **Continuous series:** $Q_1 = l + \frac{\left(\frac{N}{4} - C\right)}{f} \times i$

(3) **Upper quartile**

(i) **Discrete series:** $Q_3 = \text{size of} \left[\frac{3(n+1)}{4}\right]^{\text{th}} \text{ item}$

(ii) **Continuous series:** $Q_3 = l + \frac{\left(\frac{3N}{4} - C\right)}{f} \times i$

(4) **Decile:** Decile divide total frequencies N into ten equal parts.

$D_j = l + \frac{\frac{N \times j}{10} - C}{f} \times i$ [j = 1, 2, 3, 4, 5, 6, 7, 8, 9]

If j = 5, then $D_5 = l + \frac{\frac{N}{2} - C}{f} \times i$. Hence D_5 is also known as median.

(5) **Percentile:** Percentile divide total frequencies N into hundred equal parts

$P_k = l + \frac{\frac{N \times k}{100} - C}{f} \times i$

where k = 1, 2, 3, 4, 5,.......,99.

6. Mode.

Mode: The mode or model value of a distribution is that value of the variable for which the frequency is maximum. For continuous series, mode is calculated as, Mode $= l_1 + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times i$

Where, l_1 = The lower limit of the model class

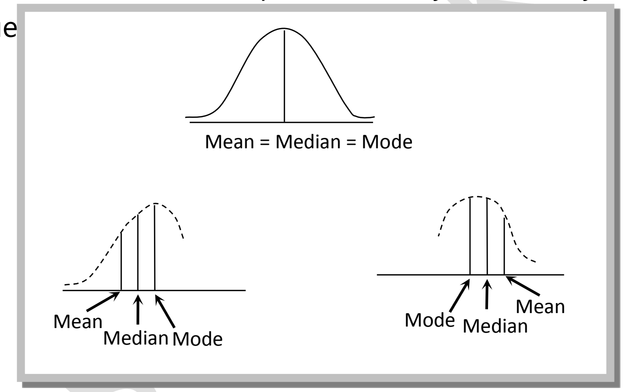
f_1 = The frequency of the model class

f_0 = The frequency of the class preceding the model class

f_2 = The frequency of the class succeeding the model class

i = The size of the model class.

Symmetric distribution: A symmetric is a symmetric distribution if the values of mean, mode and median coincide. In a symmetric distribution frequencies are symmetrically distributed on both sides of the centre point of the freque



A distribution which is not symmetric is called a skewed-distribution. In a moderately asymmetric the interval between the mean and the median is approximately one-third of the interval between the mean and the mode i.e. we have the following empirical relation between them

Mean – Mode = 3(Mean – Median) $\Rightarrow$ Mode = 3 Median – 2 Mean. It is known as Empirical relation.

Important Tips

- **Some points about arithmetic mean**
  - Of all types of averages the arithmetic mean is most commonly used average.
  - It is based upon all observations.
  - If the number of observations is very large, it is more accurate and more reliable basis for comparison.
- **Some points about geometric mean**
  - It is based on all items of the series.
  - It is most suitable for constructing index number, average ratios, percentages etc.
  - G.M. cannot be calculated if the size of any of the items is zero or negative.
- **Some points about H.M.**
  - It is based on all item of the series.
  - This is useful in problems related with rates, ratios, time etc.
  - A.M. ≥ G.M. ≥ H.M. and also $(G.M.)^2 = (A.M.)(H.M.)$
- **Some points about median**
  - It is an appropriate average in dealing with qualitative data, like intelligence, wealth etc.
  - The sum of the deviations of the items from median, ignoring algebraic signs, is less than the sum from any other point.
- **Some points about mode**
  - It is not based on all items of the series.
  - As compared to other averages mode is affected to a large extent by fluctuations of sampling,.
  - It is not suitable in a case where the relative importance of items have to be considered.

7. Pie Chart (Pie Diagram).

Here a circle is divided into a number of segments equal to the number of components in the corresponding table. Here the entire diagram looks like a pie and the components appear like slices cut from the pie. In this diagram each item has a sector whose area has the same percentage of the total area of the circle as this item has of the total of such items. For example if N be the total and n_1 is one of the components of the figure corresponding to a particular item, then the angle of the sector for this item $=\left(\frac{n_1}{N}\right)\times 360°$, as the total number of degree in the angle subtended by the whole circular arc at its centre is 360°.

8. Measure of Dispersion.

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

(1) Range

(2) Mean deviation

(3) Standard deviation

(4) Square deviation

(1) **Range:** It is the difference between the values of extreme items in a series. Range = $X_{max} - X_{min}$

The coefficient of range (scatter) $=\frac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}$.

Range is not the measure of central tendency. Range is widely used in statistical series relating to quality control in production.

(i) **Inter-quartile range:** We know that quartiles are the magnitudes of the items which divide the distribution into four equal parts. The inter-quartile range is found by taking the difference between third and first quartiles and is given by the formula

Inter-quartile range $= Q_3 - Q_1$

Where Q_1 = First quartile or lower quartile and Q_3 = Third quartile or upper quartile.

(ii) **Percentile range:** This is measured by the following formula

Percentile range $= P_{90} - P_{10}$

Where P_{90} = 90th percentile and P_{10} = 10th percentile.

Percentile range is considered better than range as well as inter-quartile range.

(iii) **Quartile deviation or semi inter-quartile range:** It is one-half of the difference between the third quartile and first quartile i.e., $\text{Q.D.} = \frac{Q_3 - Q_1}{2}$ and coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$.

Where, Q_3 is the third or upper quartile and Q_1 is the lowest or first quartile.

(2) **Mean deviation:** The arithmetic average of the deviations (all taking positive) from the mean, median or mode is known as mean deviation.

(i) **Mean deviation from ungrouped data (or individual series)**

Mean deviation $= \frac{\sum |x - M|}{n}$

Where $|x - M|$ means the modulus of the deviation of the variate from the mean (mean, median or mode). M and n is the number of terms.

(ii) **Mean deviation from continuous series:** Here first of all we find the mean from which deviation is to be taken. Then we find the deviation $dM = |x - M|$ of each variate from the mean M so obtained.

Next we multiply these deviations by the corresponding frequency and find the product f.dM and then the sum $\sum f\,dM$ of these products.

Lastly we use the formula, mean deviation $= \frac{\sum f|x - M|}{n} = \frac{\sum f\,dM}{n}$, where $n = \Sigma f$.

Important Tips

- Mean coefficient of dispersion $= \frac{\text{Mean deviation from the mean}}{\text{Mean}}$
- Median coefficient of dispersion $= \frac{\text{Mean deviation from the median}}{\text{Median}}$
- Mode coefficient of dispersion $= \frac{\text{Mean deviation from the mode}}{\text{Mode}}$
- In general, mean deviation (M.D.) always stands for mean deviation about median.

(3) **Standard deviation:** Standard deviation (or S.D.) is the square root of the arithmetic mean of the square of deviations of various values from their arithmetic mean and is generally denoted by σ read as sigma.

(i) **Coefficient of standard deviation:** To compare the dispersion of two frequency distributions the relative measure of standard deviation is computed which is known as coefficient of standard deviation and is given by

Coefficient of S.D. $= \frac{\sigma}{\bar{x}}$, where $\bar{x}$ is the A.M.

(ii) **Standard deviation from individual series**

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

where, $\bar{x}$ = The arithmetic mean of series

N = The total frequency.

(iii) **Standard deviation from continuous series**

$$\sigma = \sqrt{\frac{\Sigma f_i(x_i-\bar{x})^2}{N}}$$

where, $\bar{x}$ = Arithmetic mean of series

x_i = Mid value of the class

f_i = Frequency of the corresponding x_i

N = Σf = The total frequency

Short cut method

(i) $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

(ii) $\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$

Where, d = x – A = Deviation from the assumed mean A

f = Frequency of the item

N = Σf = Sum of frequencies

(4) **Square deviation**

(i) **Root mean square deviation**

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^2}$$

where A is any arbitrary number and S is called mean square deviation.

(ii) **Relation between S.D. and root mean square deviation:** If σ be the standard deviation and S be the root mean square deviation.

Then $S^2 = \sigma^2 + d^2$.

Obviously, S^2 will be least when d = 0 i.e. $\bar{x} = A$

Hence, mean square deviation and consequently root mean square deviation is least, if the deviations are taken from the mean.

9. Variance.

The square of standard deviation is called the variance.

Coefficient of standard deviation and variance: The coefficient of standard deviation is the ratio of the S.D. to A.M. i.e., $\frac{\sigma}{\bar{x}}$. Coefficient of variance = coefficient of S.D.$\times 100 = \frac{\sigma}{\bar{x}} \times 100$.

Variance of the combined series: If $n_1; n_2$ are the sizes, $\bar{x}_1; \bar{x}_2$ the means and $\sigma_1; \sigma_2$ the standard deviation of two series, then $\sigma^2 = \frac{1}{n_1 + n_2}[n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)]$

Where, $d_1 = \bar{x}_1 - \bar{x}$, $d_2 = \bar{x}_2 - \bar{x}$ and $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$.

Important Tips

- Range is widely used in statistical series relating to quality control in production.
- Standard deviation ≤ Range i.e., variance ≤ $(\text{Range})^2$.
- Empirical relations between measures of dispersion
  - Mean deviation $= \frac{4}{5}$(standard deviation)
  - Semi interquartile range $= \frac{2}{3}$(standard deviation)
- Semi interquartile range $= \frac{5}{6}$(mean deviation)

- For a symmetrical distribution, the following area relationship holds good

  $\overline{X} \pm \sigma$ covers 68.27% items

  $\overline{X} \pm 2\sigma$ covers 95.45% items

  $\overline{X} \pm 3\sigma$ covers 99.74% items

- S.D. of first n natural numbers is $\sqrt{\frac{n^2-1}{12}}$.
- Range is not the measure of central tendency.

---

10. Skewness.

"Skewness" measures the lack of symmetry. It is measured by $\gamma_1 = \frac{\Sigma(x_i - \mu)^3}{\{\Sigma(x_i - \mu^2)\}^{3/2}}$ and is denoted by γ_1.

The distribution is skewed if,
(i) Mean $\neq$ Median $\neq$ Mode
(ii) Quartiles are not equidistant from the median and
(iii) The frequency curve is stretched more to one side than to the other.

(1) **Distribution:** There are three types of distributions

(i) **Normal distribution:** When $\gamma_1 = 0$, the distribution is said to be normal. In this case
Mean = Median = Mode

(ii) **Positively skewed distribution:** When $\gamma_1 > 0$, the distribution is said to be positively skewed. In this case
Mean > Median > Mode

(iii) **Negative skewed distribution:** When $\gamma_1 < 0$, the distribution is said to be negatively skewed. In this case
Mean < Median < Mode

(2) **Measures of skewness**

(i) **Absolute measures of skewness:** Various measures of skewness are

(a) $S_K = M - M_d$ (b) $S_K = M - M_o$ (c) $S_k = Q_3 + Q_1 - 2M_d$

where, M_d = median, M_o = mode, M = mean

Absolute measures of skewness are not useful to compare two series, therefore relative measure of dispersion are used, as they are pure numbers.

(3) **Relative measures of skewness**

(i) **Karl Pearson's coefficient of skewness:** $S_k = \frac{M - M_o}{\sigma} = 3\frac{(M - M_d)}{\sigma}$, $-3 \le S_k \le 3$, where σ is standard deviation.

(ii) **Bowley's coefficient of skewness:** $S_k = \frac{Q_3 + Q_1 - 2M_d}{Q_3 - Q_1}$

Bowley's coefficient of skewness lies between –1 and 1.

(iii) **Kelly's coefficient of skewness:** $S_K = \frac{P_{10} + P_{90} - 2M_d}{P_{90} - P_{10}} = \frac{D_1 + D_9 - 2M_d}{D_9 - D_1}$