Measure of central tendencies :→

Mean :→ If $x_1, x_2, x_3, \dots, x_n$ are a set of n values of a variate, then the Arithmatic mean (Mean) is given by

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_i x_i}{n}$$

* In a frequency distribution, if $x_1, x_2, \dots, x_n$ represents the middle value of the class intervals having frequencies $f_1, f_2, \dots, f_n$ respectively then

$$\mu = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots f_n} = \frac{\sum_i f_i x_i}{\sum_i f_i}$$

Deviation method

$$\bar{\mu} = A + \frac{\sum f_i d_i}{\sum f_i}$$

where

$$d_i = x_i - A$$

A : Arbitary Origin

Step Deviation method

$$\mu = A + h \frac{\sum_i f_i u_i}{\sum_i f_i}$$

$$u_i = \frac{x_i - A}{h}$$

h : class interval

Obs : The algebraic sum of the deviations of all the variables from their mean is zero.

$$\sum f_i (x_i - \mu) = \sum f_i x_i - \mu \sum f_i$$

$$= \sum f_i x_i - \frac{\sum f_i x_i}{\cancel{\sum f_i}} \cancel{\sum f_i}$$

$$= 0$$

The following is the frequency distribution of weekly earning of all 509 employees of a company : find the mean

WE :	10	12	14	16	18	20	22	24	26
NofE :	3	6	10	15	24	42	75	90	79

WE :	28	30	32	34	36	38	40
N.E :	55	36	26	19	13	9	7

For a sample the formulae for mean is

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

* If we have two samples with mean $\bar{x}_1$ and $\bar{x}_2$ with number of samples n_1 and n_2 respectively then the new sample formed by combining these two samples will be

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

* Shifting the data by an amount d will aso shift the mean by d

* Scaling the data by an amount a will also scale the mean by a.

$$x \longrightarrow \bar{x}$$

$$y = ax + d \longrightarrow \bar{y} = a\bar{x} + b$$

median :→ If the values of a variable are arranged in ascending order of magnitude, median is the middle value term if the number is odd and arithmatic mean of the two middle term if the number is even. median is the value that divides the frequency in to two equal parts.

for a grouped data :→

$$\text{Median} = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

L : lower limit of median class

f : frequency of median class

C : Cummulative frequency up to class preceding the median class

h : width of median class

Mode :⇒ It is defined as that value of the variable which occurs with most frequency. value with highest frequency.

For a grouped data :→

$$\text{mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

L : lower limit of the class containing the mode

Δ_1 : excess of the modal frequency over the following class

Δ_2 : excess of the modal frequency over the preceding class

h : class - width.

Relationship between mean, median, mode :⇒

In a symmetrical distribution mean, median and mode coincide.

$$(\text{mean} - \text{mode}) = 3(\text{Mean} - \text{Median})$$

Geometric mean :→ If $x_1, x_2, \ldots, x_n$ are set of n observation, then geometric mean is defined as

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

$$\log(G.M.) = \frac{1}{n}(\log x_1 + \log x_2 + \ldots + \log x_n)$$

⇒ Let $x_1, x_2, \ldots, x_n$ be the central values with corresponding frequencies $f_1, f_2, \ldots f_n$ respectively then

$$G.M. = \left(x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_n^{f_n}\right)^{1/\alpha}$$

$$\alpha = \sum_i f_i$$

Harmonic mean

$$H.M. = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}\right)}$$

for a grouped data ;

$$H.M. = \frac{1}{\frac{1}{N}\left(\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots \frac{f_n}{x_n}\right)}$$

$$N = \sum_i f_i$$

Q An incomplete frequency distribution is given

x :	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f :	12	30	f_0 ?	65	f_1 ?	25	18

total frequency = 229 median = 46 ; find the missing frequencies median = 46 so median class is 40-50.

we know that

$$\text{median} = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

$$46 = 40 + \frac{\left(\frac{229}{2} - (12+30+f_b)\right)}{65} \times 10$$

$$\Rightarrow \frac{6 \times 65}{10} = \frac{229}{2} - (42 + f_b)$$

$$f_b = \frac{229}{2} - 42 - \frac{390}{10}$$

$$= 33.5 \approx 34$$

$$f_L = 229 - 184 = 45$$

fact Sum of deviation of $x_1, x_2, x_3, \dots, x_n$ from their mean is equal to ZERO.

Q Prove that for a given set of observations the sum of the squares of deviations is minimum when deviations are taken from mean.

$$\sum_{i=1}^{n} (x_i - A)^2 \rightarrow \text{to minimize}$$

$$x_i - A = (x_i - \mu) + (\mu - A)$$

so, $$\sum_{i=1}^{n} (x_i - A)^2 = \sum_{i=1}^{n} (x_i - \mu)^2 + 2(x_i - \mu)(\mu - A) + (\mu - A)^2$$

$$= \sum_{i=1}^{n} (x-\mu)^2 + 2(\mu-A) \underbrace{\sum_{i} (x_i-\mu)}_{0} + (\mu-A)^2 \sum_{i} 1$$

$$= \sum_{i=1}^{n} (x-\mu)^2 + 0 + (\mu-A)^2 \cdot n$$

$$\sum (x_i-A)^2 = \underbrace{\sum (x-\mu)^2}_{\text{does not depend on } A} + \underbrace{n(\mu-A)^2}_{\text{this should be minimum}}$$

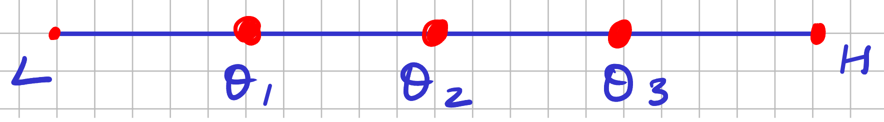
$\because$ $n(\mu-A)^2$ is +ve its minimum value will be ZERO

---

so minimum value is when $\boxed{\mu = A}$

$$\Rightarrow \boxed{A.M. \geq G.M. \geq H.M.}$$

Quartiles :→ Quartiles are those values which divides the frequency into four equal parts, when the values are arranged in ascending order of magnitude. The **lower quartile (Q_1)** is mid way between lower extreme and median. and **upper quartile (Q_3)** is midway between median and upper extreme.



$$Q_1 = L + \frac{\left(\frac{N}{4} - c\right)}{f} \times h$$

$$Q_3 = L + \frac{\left(\frac{3N}{4} - c\right)}{f} \times h$$

L: lower limit of quartile class

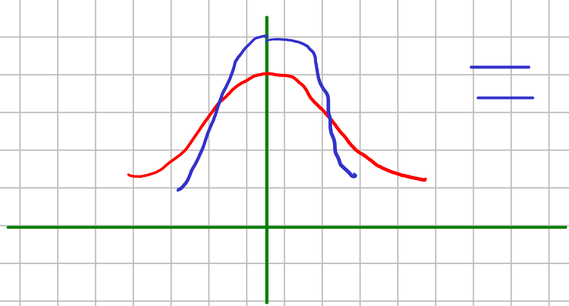
C: cummulative frequency upto the class preceding the quartile

N: Total frequency

f: width of the quartile class

Measures of dispersion :→ Dispersion describes the size of the distribution of values expected from a particular variable

It means the extent to which a variable is likely to vary about its average value.



= same mean but different dispersion.

Range : It is the simples measure of dispersion and is given by the difference of greatest and least value in the distribution.

$$range = Max. - Min.$$

Quartile Deviation: If Q_1 and Q_3 are first and third quartiles, then semi-interquartile range or quartile deviation is given by

$$Q = \frac{1}{2}(Q_3 - Q_1)$$

Mean-Deviation: → The mean deviation is the mean of the absolute differences of the values from the mean, median or mode.

$$M.D. = \frac{1}{n} \Sigma f_i |x_i - A|$$

where $A \Rightarrow$ mean, median, mode

Variance: → variance measures how far the data is spread. It is defined as the average of the squared differences from the mean.

for population;

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2$$

where $\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i$

(*) Standard deviation is square root of variance

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2} \Rightarrow$$ has same unit as that of the μ_x

Deviation method

$$\sigma^2 = \frac{\sum f_i d_i^2}{N} - \left[\frac{\sum f_i d_i}{N}\right]^2$$

Step-Deviation method

$$\sigma^2 = h^2\left[\frac{\sum f_i d_i'^2}{N} - \left\{\frac{\sum f_i d_i'}{N}\right\}^2\right]$$

here $d_i = x_i - A$ A : assumed mean

$d_i' = \frac{x_i - A}{h}$ h : class interval

$N = \sum_i f_i \Rightarrow$ Total frequency

Proof :→

$$x_i - \mu_x = (x_i - A) - (\mu_x - A)$$

$$\sum f_i (x_i - \mu_x)^2 = \sum f_i \left[(x_i - A)^2 - 2(x_i - A)(\mu_x - A) + (\mu_x - A)^2\right]$$

$$= \sum f_i d_i^2 + (\mu_x - A)^2 \sum f_i - 2\,\underline{(\mu_x - A)\sum f_i d_i}$$

$$= \sum f_i d_i^2 + \underline{(\mu_x - A)^2 \sum f_i} - 2(\mu_x - A)(\mu_x - A)\sum f_i$$

$$\mu_x = A + \frac{\sum f_i d_i}{\sum f_i} \Rightarrow (\mu_x - A)\sum f_i = \sum f_i d_i$$

$$= \sum f_i d_i^2 - (\mu_x - A)^2 \sum f_i$$

$$\sum f_i (x-\mu_x)^2 = \sum f_i d_i^2 - (\mu_x - A)^2 \sum f_i$$
$$= \sum f_i d_i^2 - \left(\frac{\sum f_i d_i}{\sum f_i}\right)^2 \sum f_i$$

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

$$\boxed{\sigma^2 = \frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2}$$

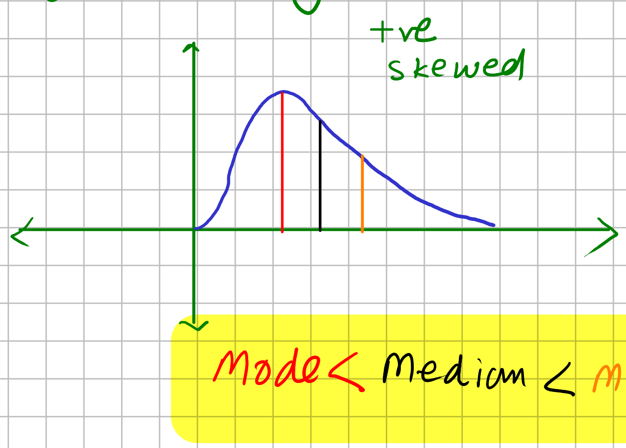
Coefficient of variation :→ It is a ratio of the standard deviation to the mean. Since this quantity does not depends upon the unit of the data this can be used to compare any two more variables that are in different units.

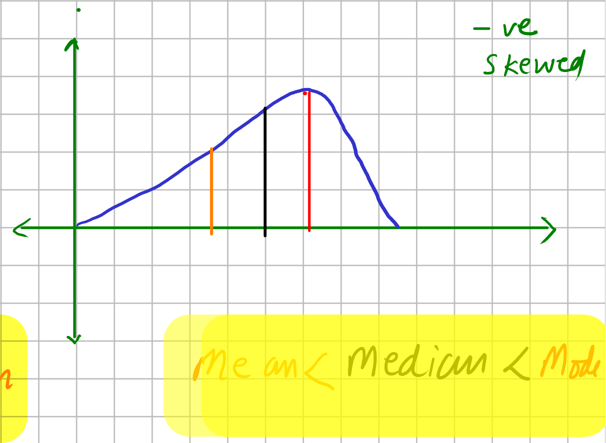
$$\boxed{COV = \frac{\sigma_x}{\mu_x}}$$

~~Et~~

* find the standard deviation of combination of two groups having different mean standard deviation and size.

SKEWNESS :→ It measures the degree of asymmetry or the departure from symmetry :





(I) Pearson's Coefficient of Skewness $= \frac{\text{mean} - \text{mode}}{\sigma}$

$= \frac{3(\text{mean} - \text{median})}{\sigma}$

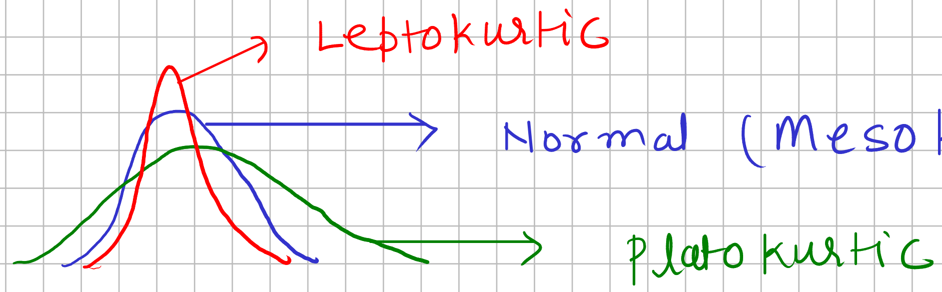
(II) Quartile Coefficient of Skewness $= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

↳ always between +1 & -1

KURTOSIS :→ Prof. Pearson termed it as "Convexity of the Curve".

It gives a measure of flatness of the distribution.

The degree of kurtosis of a distribution is measured relative to that of the ~~NORMAR~~ CURVE.



1. Pearson's Measure of kurtosis :→ using second and third central moment this

measure is defined

Here ;

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\gamma_2 = \beta_2 - 3$$

if :

1. if $\beta_2 = 3$ or $\gamma_2 = 0 \Rightarrow$ mesokurtic
2. if $\beta_2 < 3$ or $\gamma_2 < 0 \Rightarrow$ platykurtic
3. if $\beta_2 > 3$ or $\gamma_2 > 0 \Rightarrow$ leptokurtic

β and γ coefficient of skewness

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$\gamma_1 = \pm\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3}$$

Ex First four moments about the mean of a distribution are ; $\mu_1 = 0$, $\mu_2 = 2.5$, $\mu_3 = 0.7$, $\mu_4 = 18.75$.

Find coefficient of skewness and kurtosis.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(0.7)^2}{(2.5)^3} = \frac{0.49}{}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.75}{(2.5)^2} = \frac{18.75}{6.25}$$

Q The first four raw moments (moments around origin) of a distribution is given as 2, 136, 320 and 40,000. Find the coefficient of moment and kurtosis.

$$\mu_1' = 2 \qquad \mu_2' = 136 \qquad \mu_3' = 320 \qquad \mu_4' = 40{,}000$$

$$\mu_1 = \mu_1' - \mu_1' = 2-2 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 136 - (2)^2 = 132$$

$$\mu_3 = E[(x-\mu_1')^3] = E[x^3 - 3x^2\mu_1' + 3x(\mu_1')^2 - (\mu_1')^3]$$

$$= E[x^3] - 3\mu_1' E(x^2) + 3\mu_1'^2 E(x) - (\mu_1')^3 E(1)$$

$$= \mu_3' - 3\mu_1'\mu_2' + 3(\mu_1')^2\mu_1' - (\mu_1')^3$$

$$= \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

$$= 320 - 3 \times 136 \times 2 + 2(2)^3$$

$$= 320 - 816 + 16$$

$$= 320 - 800 = -480$$

$$\mu_4 = E[(x-\mu_1')^4] = E[x^4 - 4x^3\mu_1' + 6x^2(\mu_1')^2 - 4x(\mu_1')^3 + (\mu_1')^4]$$

$$= E[x^4] - 4\mu_1' E(x^3) + 6(\mu_1')^2 E(x^2) - 4(\mu_1')^3 E(x) + (\mu_1')^4$$

$$= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 4(\mu_1')^4 + (\mu_1')^4$$

$$= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$

$$= 40{,}000 - 4 \times 480 \times 2$$