C.S.E 2-1 31-5-13

Central Tendency

Formula

Population variance:

$$\sigma^2 = \frac{\Sigma (x_i - \mu)^2}{n} = \frac{\Sigma x_i^2 - n\mu^2}{n}$$

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

Population mean $= \mu = \frac{\Sigma x_i}{N}$ | $\mu = \frac{\Sigma f_i x_i}{N} = \frac{\Sigma f_i x_i}{\Sigma f_i}$

Sample variance:

$$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = \frac{\Sigma x_i^2 - n\bar{x}^2}{n-1}$$

$$s^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N-1} = \frac{\Sigma f_i x_i^2 - N\bar{x}^2}{N-1}$$

$$\boxed{N = \Sigma f_i}$$

Sample mean: $\bar{x} = \frac{\Sigma x_i}{n}$

$$\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{\Sigma f_i x_i}{N}$$

Md. Harun-Ar-Rashid.

Coefficient of variation:

Standard deviation is an absolute measure of dispersion. The corresponding relative measure is known as the coefficient of variation.

If μ is the population mean & σ is the population standard deviation, then coefficient of variation denoted by C.V is defined by

$$C.V = \frac{\sigma}{\mu} \times 100.$$

If $\bar{x}$ is the mean and s is the standard deviation of a sample data set then coefficient of variation is defined by

$$C.V = \frac{s}{\bar{x}} \times 100.$$

C.V is a pure number and expressed as a percentage. It is useful in comparing the variability of two or more data sets, especially if they are expressed in different units of measurement.

Among two or more data sets, the set for which coefficient of variation is greater is said to be more variable or conversely less uniform, less stable or less consistent, less homogenous. on the other hand, the series for which coefficient of variation is less is said to be less variability or more consistent, more uniform, more stable or more homogenous

lives of two models of refrigera in a recent survey were found as follows:

life (no. of years)	Model A	Model B
0 - 2	5	2
2-4	16	7
4-6	13	12
6-8	7	19
8-10	5	9
10-12	4	1

What is the average life of each model of these refrigerators?

Which of the two models shows more uniformity?

A person wants to buy a new refrigerator, which one will he prefer?

Class interval	mid points x_i	Model A			Model B		
		f_i	$f_i x_i$	$f_i x_i^2$	f_i	$f_i x_i$	$f_i x_i^2$
0-2							
2-4							
4-6							
6-8							
8-10							
10-12							
Total		$\Sigma f_i =$	$\Sigma f_i x_i$	$\Sigma f_i x_i^2$	Σf_i	$\Sigma f_i x_i$	$\Sigma f_i x_i^2$

Model A — Model B

Average life times = Arithmetic mean.

Model A

$$\bar{x}_A = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$= \frac{\Sigma f_i x_i}{N}$$

$$= \frac{256}{50}$$

$$= 5.12 \text{ years}$$

Model B

$$\bar{x}_B = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$= \frac{304}{50}$$

$$= 6.16 \text{ years}$$

Model A

$$s_A^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N-1}$$

$$= \frac{\Sigma f_i x_i^2 - N\bar{x}^2}{N-1}$$

$$= \frac{1706 - 50 \times (5.12)^2}{50-1}$$

$$= 8.07$$

$$s_A = 2.84 \text{ years}$$

$$C.V(A) = \frac{s_A}{\bar{x}_A} \times 100$$

$$= \frac{2.84}{5.12} \times 100$$

$$= 55.47\%$$

Model B

$$s_B^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N-1}$$

$$= \frac{\Sigma f_i x_i^2 - N\bar{x}^2}{49}$$

$$= \frac{2146 - 50 \times (6.16)^2}{49}$$

$$= 5.08$$

$$s_B = 2.25 \text{ years}$$

$$C.V(B) = \frac{2.25}{6.16} \times 100$$

$$= 36.53\%$$

Comment:

(i) Average lifetimes of refrigerators of model A is 5.12 years, while of model B is 6.16 years.

(ii) Since coefficient of variation is less for model B, Hence refrigerators of model B show greater uniformity as per the lifetime of the refrigerators.

(iii) Due to the greater uniformity in lifetime, the person will prefer model B.

\# The following are some of the particular of the distribution of weights and heights of 100 boys in a class

	Weights	Heights
Mean	60 kgs	65 inches
Standard deviation	7 kgs	4 inches

Which distribution shows greater variability.

Solution: For determining variability between the two distributions, we have to compute coefficient of variations of weights and heights of the boys.

$$c.v\,(\text{weight}) = \frac{\sigma}{\mu} \times 100 = \frac{7}{60} \times 100 = 11.67\%$$

$$c.v\,(\text{Heights}) = \frac{\sigma}{\mu} \times 100 = \frac{4}{65} \times 100 = 6.15\%$$

Comment: Since coefficient of variation is higher for weights, Hence the distribution of weights shows greater variability than heights.

Exercise: In two factories A and B engaged in the same industry, the average weekly wages and standard deviations are as follows:

	Factory A	Factory B
Average weekly wages		

	Factory A	Factory B
Average weekly wages (in taka) :	860	900
Standard deviation of wages (in taka)	50	80

(i) Which factory shows greater variability in the distribution of wages?

(ii) A person got jobs in both the factories, which factory he will prefer to join?

Solution: For comparing the variability of wages, of the two factories, we have to compute coefficient of variation:

Factory A

$$C.V = \frac{\sigma}{\mu} \times 100$$
$$= \frac{50}{860} \times 100$$
$$= 5.81\%$$

Factory B

$$C.V = \frac{\sigma}{\mu} \times 100$$
$$= \frac{80}{900} \times 100$$
$$= 8.89\%$$

*Comment:

The coefficient of variation is greater for factory B. Hence factory B shows greater variability in the distribution of wages. The person will join factory A, since the coefficient of variation of the factory A is less than the factory B, although the average weekly wages of factory B is more than A.

Two cricketers scored the following runs in randomly selected 10 - one day matches:

Player A	42	32	40	45	17	83	59	64	76	72
Player B	95	03	28	70	31	14	82	0	59	108

(i) Who is the better run getter?

(ii) " " " consistent player?

(iii) A prize is given to the best player, who will get the prize?

Solution:

In order to find out who is better run getter, we will compare the average runs scored and to find out who is more consistent, we will compare the coefficient of variation.

Cricketer A		Cricketer B	
x	x^2	x	x^2
$\Sigma x =$	$\Sigma x^2 =$	$\Sigma x =$	Σx^2

Cricketer A.

$\bar{x}_A = \frac{\Sigma x}{n}$

$= \frac{530}{10}$

$= 53$

Cricketer B

$\bar{x}_B = \frac{\Sigma x}{n}$

$= \frac{490}{10}$

$= 49$

(i) Since the average score for the player A is higher than player B, Hence A is a better run getter. However

$s_A^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$

$= \frac{\Sigma x_i^2 - n\bar{x}^2}{n-1}$

$= \frac{32127 - 10\times(53)^2}{10-1}$

$= 448.56$

$\boxed{S_A = 21.18}$

$s_B^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$

$= \frac{\Sigma x_i^2 - n\bar{x}^2}{10-1}$

$= \frac{37744 - 10\times(49)^2}{9}$

$= 1526$

$\boxed{S_B = 39.06}$

$\therefore C.V(A) = \frac{s}{\bar{x}} \times 100$

$= \frac{21.18}{53} \times 100$

$= 39.96\%$

$C.V(B) = \frac{39.06}{49} \times 100$

$= 79.71\%$

(ii) The coefficient of variation for player A is less than player B; Hence player A is more consistent.

(iii) The player A will get the prize.

Advantages of coefficient of variation over standard deviation:

Standard deviation and the original variable have the same unit of measurements. So standard deviation cannot be used to compare the variability of two or more distributions measured in different units. But coefficient of variation is a pure number. That is coefficient of variation is independent of the unit of measurement of the original variables. Therefore, coefficient of variation can be successfully used to compare the variability of two or more distribution in different units of measurment. The coefficient of variation is useful also in following cases:

(I) The data are in different units.

(II) The data are in same units but the means are far apart.

(III) when the data set involves all or nearly all positive values.

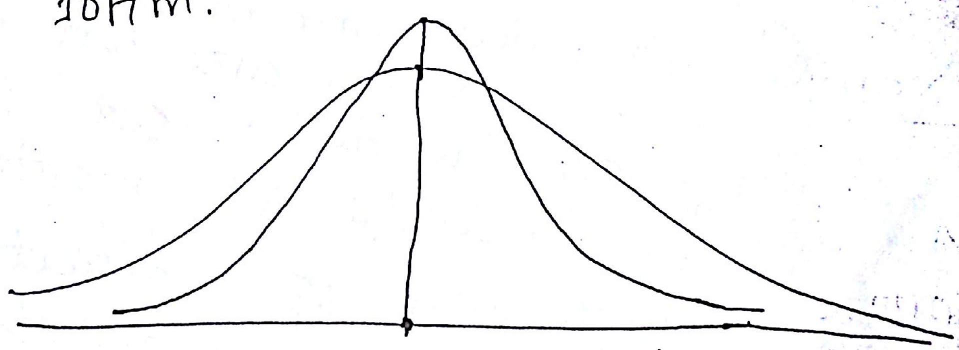
Hence coefficient of variation is sometimes better than standard deviation as a measure of dispersion.

Skewness

Skewness means the lack of symmetry of a distribution. That is when a distribution is not symmetrical, it is called skewed distribution. A distribution may be symmetrical, positively skewed or negatively skewed.

Symmetrical or zero-skewed distribution:

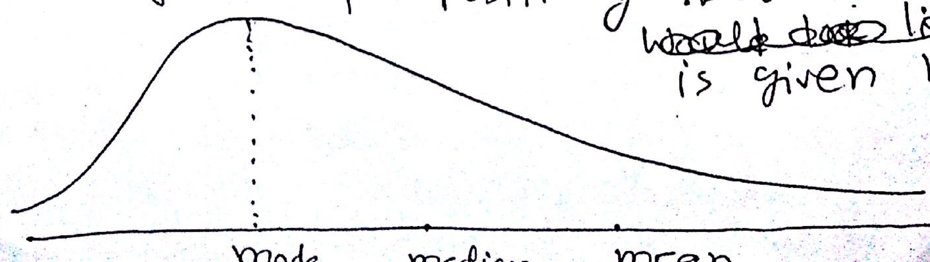
A distribution is symmetric if this curve is folded at the center, two sides will coincide. In this case the values of mean, median and mode of the distribution are equal. Normal distribution is an example of a symmetrical distribution. The diagram of symmetrical distribution will take the following form.



mean = median = mode

Positively skewed distribution:

The diagram of positively skewed distribution is given below.



mode median mean

~~In this distribution, the long tail to the right indicates the presence of extreme values at the positive end of the distribution~~

A distribution is called negatively skewed if a greater proportion of the observations lie to the right of the peak value. In this case

mean > median > mode.

<u>Example:</u> Family size, female age at marriage.

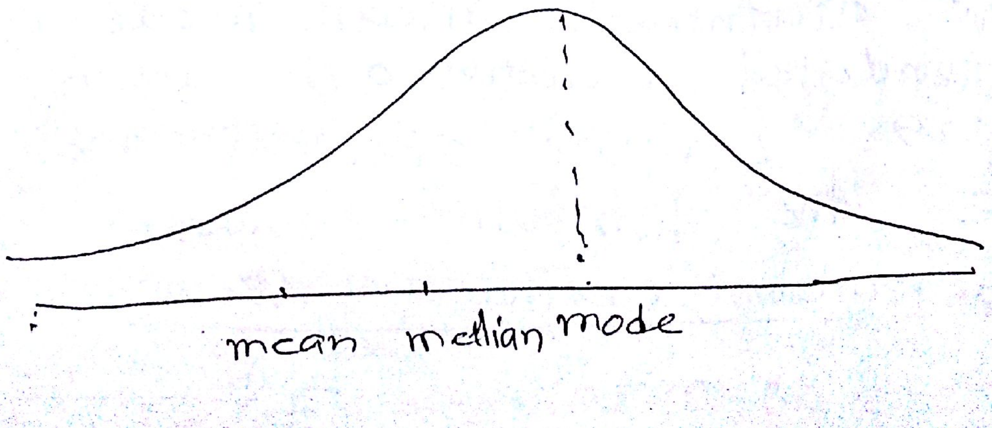
<u>Negatively skewed distribution:</u>

A distribution is called negatively skewed if a greater proportion of the observations lie to the left of the peak value. In this case

mode > median > mean.

<u>Example:</u> (i) Reaction times for an experiment.
(ii) Daily maximum temparature for winter months.

A diagram of a negatively skewed curve is given below:



Measures of skewness:

Pearson's coefficient of skewness:

$$S_K(P) = \frac{\text{mean} - \text{mode}}{\text{Standard deviation}}$$

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

If

$S_K(P) = 0$, distribution is symmetrical (no skewness)

$S_K(P) > 0$, distribution is positively skewed

$S_K(P) < 0$, distribution is negatively skewed.

Bowley's coefficient of skewness:

$$S_K(B) = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

if $S_K(B) = 0$, distribution is symmetrical (no skewness)

$S_K(B) > 0$, distribution is positively skewed

$S_K(B) < 0$, distribution is negatively skewed.

The arithmetic mean, mode and standard deviation are 37.70, 36.67 and 8.29. compute skewness and comment on the distribution:

~~Sol~~ Pearson's coefficient of skewness:

$$= \frac{\text{mean} - \text{mode}}{\text{Standard deviation}}$$

$$= \frac{37.25 - 36.67}{8.29} = 0.07$$

Comment: The value of skewness is only 0.07. Hence the distribution is slightly positively skewed.

#

The mean and median wages per day of of a worker of an industry are TK. 157 and TK. 160. Suppose the standard deviation of wages are TK 50. Calculate the coefficient of skewness and comment.

Ans. $S_K(P) = -0.18$, negatively skewed distribution.

#

The first and third Quartile of the distribution are 10 and 25. The median of the distribution is 20. find its coefficient of skewness:

Bowley's coefficient of skewness:

$$S_K(B) = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

$$= \frac{(25-20)-(20-10)}{25-10}$$

$$= \frac{-5}{15} = -0.33.$$

The distribution is negatively skewed.

The measure of skewness of a distribution is 0.3. The mode and the median are 50 and 55. Find the mean and, standard deviation of the distribution.

Ans. 57.5 & 25