Question: Given the following distribution, calculate the sample variance.

Age in Years	10 <20	20 <30	30 <40	40 <50
Frequency	10	6	20	15

4.4 Standard Deviation

The standard deviation is a positive square root of the variance. Because dealing with original units of measurement is easier than dealing with the square of the units, analysts usually use the standard deviation to measure variation in a population or sample.

Population Standard Deviation

For a set of values, the formula for computing the population standard deviation is as follows:

Population Standard deviation= $\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$

Where σ = the population standard deviation

μ = the arithmetic mean of the population

x= an observation of the population

N= the population size

Alternatively, the following formula is used:

$$\sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{N}}$$

Example: Find the population standard deviation of the following values:

2, 4, 9, 13 and 17.

Solution: $\mu = 45 \div 5 = 9$ and N=5

x	$(x-\mu)$	$(x-\mu)^2$
2	-7	49
4	-5	25
9	0	0
13	4	16
17	8	64
Total		154

The populationσ= √154=12.41

Question: Calculate the population standard deviation for the following data:

4, 7, 5, 9, 10, 8, 13

Sample Standard Deviation

For a set of values, the formula for computing the sample standard deviation is as follows:

Sample Standard deviation= s= $\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

Where s= the sample standard deviation

x□ = the arithmetic mean of the sample.

N =the sample size

X = an observation of the sample

Example: Find the sample standard deviation of the following data and interpret your findings:

2, 3, 5 and 10

Solution: : $\bar{x} = 20 \div 4 = 5$

x	$(x - \bar{x})$	$(x - \bar{x})^2$
2	-3	9
3	-2	4
5	0	0
10	5	25
Total		38

The sample standard deviation is therefore s= $\sqrt{\frac{38}{4-1}}$

= 3.56

Interpretation: Each value in the distribution is on average about 3.56 units away from the common mean.

Question: Calculate the sample standard deviation of the following data:

23, 43, 14, 26, 18, 16, 25

Population Standard Deviation for a Simple Frequency Distribution

The formula for computing the population standard deviation is shown below:

$$\sigma = \sqrt{\frac{\sum fx^2 - \frac{(\sum fx)^2}{\sum f}}{\sum f}} \text{ or}$$

Or $\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$ or $\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\bar{x}\right)^2}$

Where σ =the population standard deviation

x =an observation of the population

f = the x class frequency

$\sum f$ =the number of observations (total frequency)

Example: Find the sample standard deviation of the following distribution

x	1	2	3	4	5
f	2	5	3	12	4

Solution:

x	f	x^2	fx	fx^2
1`	2	1	2	2
2	5	4	10	20
3	3	9	9	27
4	1	16	4	16
5	4	25	20	100
Totals	15		45	165

$$\sigma = \sqrt{\frac{165 - \frac{45^2}{15}}{15}}$$

=1.41

Question: Given the following data, calculate the population standard deviation.

x	0	1	2	3
f	4	12	10	5

Sample Standard Deviation for a Simple Frequency Distribution

For a simple frequency distribution, the formula for computing the sample standard deviation is as follows:

Sample standard deviation

$$s = \sqrt{\frac{\sum fx^2 - \frac{\left(\sum fx\right)^2}{\sum f}}{\sum f - 1}}$$

Where s=the sample standard deviation

x =an observation of the sample

f = the x class frequency

$\sum f$ =the number of observations (sample size or total frequency)

Example: Find the sample standard deviation of the following distribution

x	1	2	3	4	5
f	2	5	3	12	4

Solution:

x	f	x^2	fx	fx^2
1`	2	1	2	2
2	5	4	10	20
3	3	9	9	27
4	1	16	4	16
5	4	25	20	100
Totals	15		45	165

Therefore, $s = \sqrt{\dfrac{165 - \dfrac{45^2}{15}}{15-1}}$

=1.46

This means that each value is on average about 1.46 units away from the common mean

Question: Calculate the sample standard deviation from the following data:

x	1	2	3	4	5
f	3	5	10	2	1

Population Standard Deviation for Grouped Data

The same formula is used in calculating the population standard deviation for grouped data. Thus, we have

$$\sigma = \sqrt{\frac{\sum fx^2 - \frac{\left(\sum fx\right)^2}{\sum f}}{\sum f}}$$

Example: The data below relate to the number of successful sales made by salesmen employed by Mama's Kitchen in a particular period. Calculate the population standard deviation of the number of sales.

Number of sales (x)	0 to 4	5 to 9	10 to 14	15 to 19	20 to 24	25 to 29
Number of salesmen (f)	1	14	23	21	15	6

Solution:

Number of sales	f	Mid-point (x)	fx	fx^2
0 to 4	1	2	2	4
5 to 9	14	7	98	686
10 to 14	23	12	276	3312
15 to 19	21	17	357	6069
20 to 24	15	22	330	7260
25 to 29	6	27	162	4374
Total	80	87	1225	21705

$$\sigma = \sqrt{\frac{21705 - \frac{1225}{80}}{80}}$$

= 6.1 sales

Question: A firm tabulated the ages of its regular customers to give the following:

Age in years	15 <20	20 < 25	25 <30	30 <35	35 <40

Percentage of customers	10	20	25	15	30

Estimate the standard deviation age.

Sample Standard Deviation for Grouped Data

If the data of interest are in a grouped form, the formula for finding the sample standard deviation is the same as the one used for the simple frequency distribution. However, in a grouped frequency distribution, the mid- point of each class is used for the computation.

Example 1: A sample of semi-monthly amounts invested in a company by its employees was organized into a frequency distribution for further study.

Estimate the sample standard deviation of the data and interpret your answer.

Amount invested ($)	30 and under 35	35 and under 40	40 and under 45	45 and under 50	50 and under 55	55 and under 60	60 and under 65	65 and under 70
Number of employees (*f*)	3	7	11	22	40	24	9	4

Solution:

Amount invested ($)	**Number of employees (frequency)**	**Mid-Point (x)**	x^2	Fx	fx^2
30 and under	3	32.5	1,056.25	97.5	3168.75

35					
35 and under 40	7	37.5	1,406.25	262.5	9843.75
40 and under 45	11	42.5	1,806.25	467.5	39,737.5
45 and under 50	22	47.5	2,256.25	1,045	49,637.5
50 and under 55	40	52.5	2,756.25	2,100	110,250
55 and under 60	24	57.5	3,306.25	1,380	79,350
60 and under 65	9	62.5	3,906.25	562.5	35,156.25
65 and under 70	4	67.5	4,556.25	270	18225
Total	120	400	21,050.00	6185	345368.75

Sample standard deviation

$$s = \sqrt{\frac{345{,}368.75 - \frac{(61`85)^2}{120}}{120 - 1}}$$

$s = \$14.95$

Interpretation: Each amount invested by the workers is about $14.95 away from the common mean amount.

Example 2: Estimate the sample standard deviation of the following distribution and interpret your result.

Class	1 up to 5	6 up to 10	11 up to 15	16 up to 20	21 up to 25
Frequency	2	7	12	6	3

Solution:

Class	Frequency	Mid-Point	x^2	fx	fx^2
1 up to 5	2	3	9	6	18
6 up to 10	7	8	64	56	448
11 up to 15	12	13	169	156	2028
16 up to 20	6	18	324	108	1944
21 up to 25	3	23	529	69	1587
Total	30	65	1095	395	6025

Standard deviation = $s = \sqrt{\dfrac{6025 - \dfrac{(395)^2}{30}}{30-1}}$

$s = \sqrt{\dfrac{6025 - 5200.83}{29}}$

$= \sqrt{824.17} = 28.71$

Question: Given the following data, calculate the standard deviation of the distribution:

Cost	5 and under 10	10 and under 15	15 and under 20	20 and under 25	25 and under 30
Frequency	1	5	10	6	1

4.6 Relative Dispersion

For distributions having the same arithmetic mean, the distribution with the largest standard deviation has the greatest relative spread. However, when two or more distributions have different means, the relative spread cannot be determined by merely comparing standard deviations. In that case, the coefficient of variation is used.

Coefficient of Variation

Coefficient of variation is simply the standard deviation as a percentage of the arithmetic mean. It is used in measuring the relative variation (dispersion) for distributions with different means. When the coefficients of variation for two or more distributions are compared, the distribution with the largest coefficient of variation is said to have the largest relative spread (dispersion).

The formula for computing the coefficient of variation for a population data is as follows:

$$CV= \frac{\sigma}{\mu}\times 100$$

For sample data, the formula is as follows:

$$CV=\frac{s}{\bar{x}}\times 100$$

In finance, the coefficient of variation measures the relative risk of a stock portfolio.

Example 1: XYZ Incorporated has 2 stock portfolio, A and B. Portfolio A has a collection of stocks that average a 12% return with a standard deviation of 3% and portfolio B has an average return of 6% with a standard deviation of 2%. Compute the coefficient of variation and draw your conclusion.

Solution:

For Portfolio A,

$$CV=\frac{3}{12}\times 100 = 25\%$$

For portfolio B,

$$CV= \frac{2}{6}\times 100 = 33.33\%$$

Conclusion: Even though the standard deviation of portfolio B is lower, it has a greater relative dispersion than portfolio A because 33% is greater than 25%

The coefficient of variation can be used to compare the relative spread of other distributions.

Example 2: For a sample of executives, the mean income is $500,000 and the standard deviation is $50,000. For a sample of unskilled workers, the mean income is $22,000 and the standard deviation is $2,200. Compute the coefficient of variation and interpret your result.

Solution: For the executives,

$$CV = \frac{50,000}{500,000} \times 100 = 10\%$$

For the unskilled workers,

$$CV = \frac{2,200}{22,000} \times 100 = 10\%$$

Interpretation: There is no difference in the relative variation (dispersion) of the two groups because they both have a coefficient of variation of 10%

It must be noted that the coefficient of variation can be used to measure relative variation even if the units of measurements for 2 or more distributions are different.

Questions:

1**.** Over a period of 3 months, the daily number of components produced by two comparable machines was measured giving the following statistics:

Machine A: Arithmetic mean= 242.8 and standard deviation =20.5

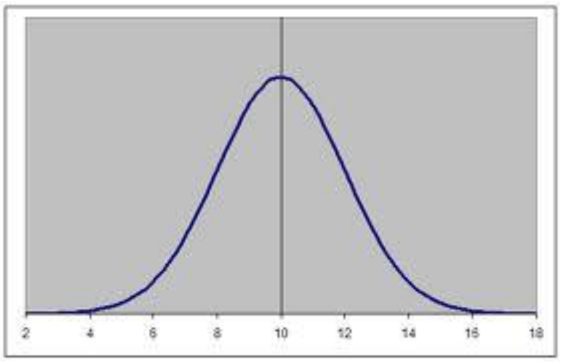
Machine B: Arithmetic mean= 281.3 and standard deviation =23.0

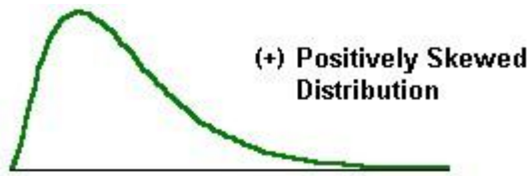
2. A study of the test score for an in-plant course in management principles and the years of service of the employees enrolled in the course resulted in these statistics: The mean test score was 200 and the standard deviation was 40. The mean number of years of service was 20 years and the standard deviation was 2 years. Compare the relative dispersion of the two distributions using the coefficient of variation and interpret your result.

4.7 Coefficient of Skewness

Another important characteristic that can be measured is the degree of skewness of a distribution. A distribution can either be positively (or right) skewed, negatively (or left) skewed or symmetrical. If a distribution is symmetrical, it has no skewness. Consider the following distributions:

Symmetrical Distribution (No Skewness)





Coefficient of skewness is used in measuring the skewness of a distribution. The formula is as follows:

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

Example 1: The length of stay on the cancer floor of Korle-bu Teaching hospital was organized into a frequency distribution. The mean length of stay was 28 days, the median 25 days and the modal length 28 days. The standard deviation was computed to be 4.2. Determine the coefficient of skewness and interpret your findings.

Solution:

$$S_k = \frac{3(28-25)}{4.2}$$

$$= 2.14$$

Interpretation: There is a substantial amount of positive skewness

Example 2: For a group of executives in the electronics industry, the mean salary is \$542,000 and the median is \$400,000. The standard deviation is \$ 448,500. Determine the coefficient of skewness. And comment on your result.

$$S_k = \frac{3(542{,}000 - 400{,}000)}{448{,}500}$$

$$= 0.32.$$

Comment: The distribution is positively skewed.

Example 3: For a particular distribution of wages, the mean was computed to be GH¢25,000, the median GH¢25,000 and the mode GH¢25,000. The standard deviation was GH¢1,000. Determine the coefficient of skewness and comment on your result.

.

Solution:

$$S_k = \frac{3(25{,}000 - 25{,}000)}{1{,}000} = 0$$

Comment: There is no skewness and so the distribution is symmetrical.

Example 4: Given the following population data: 2, 5, 3, 4, and 16, compute the coefficient of skewness.

Solution:

Mean= $\frac{\sum x}{n} = \frac{30}{5} = 6$

Median= 4 since it is the middle value in the data.

You need to find the population standard deviation since it is a population data. So

$\sigma = \sqrt{\frac{130}{5}}$ =5.1

Therefore $S_k = \frac{3(6-4)}{5.1`}$ =1.18 (2D)

Conclusion: The distribution is positively skewed.

Question: Given the following distribution

Amount (GH¢)	Male	Female
10 and under 12	5	8
12 and under 14	8	10
14 and under 16	7	6
16 and under 18	4	14
18 and under 20	6	2

(a) Estimate the means for the two distributions.

(b) Estimate the medians for the two distributions.

(c) Estimate the standard deviations.

(d) Calculate the coefficient of skewness and comment on your results.

(e) Calculate the coefficient of variation and comment on your results.