1. Introduction

Welcome to the series of E-learning modules on Weighted Mean, Truncated Mean and Mid-Range. In this module, we are going to cover the definition, types and application of weighted mean, truncated mean and mid-range.

By the end of this session, you will be able to:

- Explain the concept of weighted mean, truncated mean and mid-range
- Explain the types of weighted mean, truncated mean and mid-range
- Discuss application of weighted mean, truncated mean and mid-range

A data set has various data points and while analyzing these data points we use various statistical measures to provide a meaningful interpretation of the data set. The measures of central tendency namely the mean, median and mode are one of the statistical tools that we use for analyzing the data before arriving at any interpretation.

Means are mathematical formulations used to characterize the central tendency of a set of numbers. Most people are familiar with the "arithmetic mean", which is also commonly called an average.
While calculating the measures of mean, equal importance is given to all the data points in the data. However, there are cases where not all the items are of equal importance. The importance of the data point is by its nature.

In other words, some items of a series are important as compared to the other items in the same series. In such cases, it becomes important to assign different weights to different items.
The weighted mean can be used to calculate an average that takes into account the importance of each value with respect to the overall total.

There are situations where we see some data points are important and the relative importance is by nature of the contribution of that data point to the analysis of the data.

Let us take an example to get an idea of the change in the cost of living of a certain group of people. A simple mean of the prices of the commodities consumed by these people will not be an appropriate tool for measuring average price, as not all the commodities may be of equal importance. For example wheat, rice and pulses may be more important when compared with cigarettes, tea and luxury items.

Weighted mean is similar to an arithmetic mean (the most common type of average) where instead of each of the data points contributing equally to the final average, some data points contribute more than others do. The idea of weighted mean plays an important role in descriptive statistics and occurs in a more general form in several areas of mathematics.

Weighted mean is a mean that is computed with extra weight given to one or more elements of the sample. Thus, arithmetic mean computed by considering relative importance of each items is called weighted arithmetic mean. To give due importance to each item under consideration, we assign number called weight to each item in proportion to its relative importance.

The term ‘weighted average’ usually refers to a weighted arithmetic mean, but weighted versions of other means can also be calculated. Such as the weighted geometric mean and the weighted harmonic mean.

2. Weighted Arithmetic Mean and Weighted Geometric Mean

Let us discuss about the Weighted Arithmetic Mean.
Weighted Arithmetic Mean is computed by using the following formula:
Weighted arithmetic mean is equal to summation of the product of the weights and the variable divided by summation of weights.
Where, Xw stands for weighted arithmetic mean, x stands for values of the items, W stands for weight of the item and f stands for frequency.

In case of ungrouped data where weights are involved, our approach for calculating arithmetic mean will be different from the approach used earlier. The steps used are:

1. Multiply the weights by the variables and obtain the total
2. Divide the total by the sum of the weights
3. In case of frequency distribution, we multiply the weights, frequency and the variables

Suppose a student has secured the following marks in three tests: mid-term test 30 marks, laboratory 25 marks and final 20 marks.

Figure 1

Test	Marks
Mid-term	30
Laboratory	25
Final	20

Then, the simple arithmetic mean will be 30 plus 25 plus 20 divided by 3 is equal to 25.

Figure 2

Test	Relative weights (w)	Marks (x)	wx
Mid-term	2	30	60
Laboratory	3	25	75
Final	5	20	100
Total	10	75	235

However, this will be wrong if the three tests carry different weights based on their relative importance. Assuming that the weights assigned to the three tests are mid-term test 2 points, laboratory 3 points and final 5 points. Then, on this basis, we can now calculate a weighted mean as shown below in the table. Where, the first column indicates the type of the test, the second column we enter the relative weights, column three shows the marks and column four shows the product of the marks and the relative weights. Now, take the total of the weights, which is equal to 10 in this case, and the total of the fourth column that is 235. Now, let us substitute the values in the equation we get 60 plus 75 plus 100 divided by 2 plus 3 plus 5 is equal to 23.5 marks.
It is seen that the weighted mean gives more realistic picture than the simple or unweighted mean.

The problem that arises while using the weighted mean is regarding selection of weights. Weights may be actual or arbitrary. Actual weights are weights rightly assigned to the variable and are the right choice if available. In the absence of actual weights, we shall use imaginary weights called arbitrary weights. Sometimes, arbitrary weights lead to errors in calculation, but it is better than no weights. It is also found that if the weights are logically assigned keeping the process in mind then, the error involved will be so small that it can be easily over looked.

Let us discuss the relationship between Simple and Weighted Arithmetic Mean.

1. Simple arithmetic mean shall be equal to the weighted arithmetic mean if the weights are equal.
2. Simple arithmetic mean shall be less than the weighted arithmetic mean if and only if greater weights are assigned to greater values and smaller weights are assigned to smaller values.
3. Simple arithmetic mean is greater than the weighted arithmetic mean if and only if smaller weights are attached to the higher values and greater weights are attached to smaller values.

Weighted Arithmetic Means are useful in problems relating to:

- Construction of Index numbers
- Standardized birth and death rates

Now, let us explain the Weighted Geometric Mean.
Geometric mean is the 'nth' root of the product of 'n' items of a series.
In case of ungrouped data where weights are involved, our approach for calculating geometric mean will be different from one used earlier. The steps used are:
1. Take the logarithm of the variables
2. Multiply the weights with the logarithm of the variables and obtain the total
3. Divide the total by the sum of the weights
4. In case of frequency distribution, we multiply the weights, frequency and the logarithms of the variables

To calculate the weighted geometric mean, we use the following formula:
Weighted geometric mean is equal to the antilog of summation of the product of weights and the logarithms of X divided by the summation of weights. We use the logarithms to calculate the geometric mean as it is difficult to find the nth root.

Let us take an example to understand the weighted geometric mean.

Find the weighted geometric mean for the following data:

Figure 3

Group	Index Number	Weights
Food	260	46
Fuel & Lighting	180	10
Clothing	220	8
House Rent	230	20
Education	120	12
Miscellaneous	200	4

In this table, we have taken the given data in the first three columns. We have calculated the logarithm of the variable X in column four, and taken the product of the weights and the logarithms in column 5. We have also taken the total of the weights, which is equal to 100, and the total of the product is equal to 233.7706.

Figure 4

Group	Index Number	Weights	Log x	wlogx
Food	260	46	2.4150	111.0900
Fuel & Lighting	180	10	2.2553	22.5530
Clothing	220	8	2.3424	18.7392
House Rent	230	20	2.3617	47.2340
Education	120	12	2.0792	24.9504
Miscellaneous	200	4	2.3010	9.2040
Total		Σw=100		Σwlogx= 233.7706

Calculating the weighted geometric mean by using the formula Weighted geometric mean is equal to the antilog of summation of the product of weights and the logarithms of X divided by the summation of weights which is equal to anti log of 233.7706 divided by 100 is equal to anti log of 2.3377 is equal to 217.6.

3. Weighted Harmonic Mean

Now, let us explain the Weighted Harmonic Mean.
Harmonic mean is quotient of “number of the given values” and “sum of the reciprocals of the given values”. Harmonic mean is another measure of central tendency and it is based on mathematic footing like arithmetic mean and geometric mean.

In case of ungrouped data where weights are involved, our approach for calculating Harmonic mean will be different from one used earlier. The steps used are:

1. Take the reciprocals of the variables
2. Multiply the weights by the reciprocals of the variables
3. Divide the total of the weights by the total of the product of the reciprocals and the weights.
4. In case of frequency distribution or grouped data, we multiply the weights, frequency and the reciprocals of the variables.

To calculate the weighted Harmonic mean, we use the following formula:
The weighted Harmonic mean is equal to summation of the weights divided by summation of the product of the reciprocals and the weights.

Let us take an example to understand the weighted Harmonic mean:
A cyclist covers his first five kilometers at an average speed of 10Km.p.h, another three Km at 8Km.p.h and the last two Km at 5Km.p.h. Find the average speed of the entire journey and verify your answer.

To find the solution, we will use the weighted harmonic mean.
The various speeds are 10, 8 and 5. Therefore, their reciprocals are one by ten, one by eight and one by five. The weights are the distance covered. In this case, it is 5, 3 and 2. By multiplying the reciprocal and the weights we get ½ plus 3/8 plus 2/5 which is equal to 51/40. The total of the weights is equal to 10. Thus, the weighted harmonic mean is equal to 10 divided by 51/40 is equal to 10 into 40 by 51 is equal to 7.84Km.p.h the average speed for the entire journey.

Figure 5

Speed	**Reciprocal (x)**	**Weights (w)**	**w/x**
10	1/10	5	½
8	1/8	3	3/8
5	1/5	2	2/5
Total		10	51/40

Let us list the advantages of weighted mean.

- Realistic: Weighted averages are more realistic. A standard average assumes that everything is created equal, but in the real world, this is not the case. Therefore, by assigning different weights to things with different values, you can come up with a more realistic average.
- Flexibility: Weighted averages are flexible. You can multiply the numbers in your series by any weight that you want to show. This is advantageous because it lets statisticians adjust numbers to fit the exact situation.
- Stock Indexing: Stock indexes often use a price weight to determine the value of the index as a whole. This is advantageous because it keeps lower-price stocks from having a disproportionate effect. By making the higher stocks, it affects the value of the average more than the lower stocks do. It is possible to create a more accurate picture of the value of the index as a whole.

Let us take a real life situation to understand the weighted mean in detail.
Assume you are comparing two basketball players where the first player made 10 shots in one game, 20 shots in another and 15 shots in a third. His average shots per game are 15. The second basketball player made 5 shots in his first game, 10 in second and 15 in his third. He made an average of 10 shots per game.

Figure 6

Game	Player 1	Weights	wx	Player 2	weights	wx
1	10	1	10	5	2	10
2	20	1	20	10	2	20
3	15	1	15	15	2	30
Total	45		30			60

However, assume the first basketball player is in his fourth season and the second basketball player is still in his first. This experience difference needs to be compensated. So assume in every game the first basketball player makes a weight of 1 and the second player makes a weight of 2 to compensate for the fact that he has not been playing as long as the first player.

This dramatically changes the results. The first basketball player's weighted average is still 45, because every number was multiplied by 1. The second's, however is:
$(5 \times 2) + (10 \times 2) + (15 \times 2) = 60$
$60/3=20$
The second player's weighted average is now 20, which more accurately reflects his skill level.

4. Truncated Mean

Let us explain the Truncated Mean.
A truncated mean or trimmed mean is a statistical measure of central tendency, much like the mean and median.
It involves the calculation of the mean after discarding given parts of a probability distribution or sample at the high and low end, and typically discarding an equal amount of both.

Truncated Mean is a method of averaging that removes a small percentage of the largest and smallest values before calculating the mean. After removing the specified observations, the trimmed mean is found using an arithmetic averaging formula.

For most statistical applications, 5 to 25 percent of the ends are discarded.
In some regions of Central Europe, it is also known as a Windsor mean.

A trimmed mean is stated as a mean trimmed by X%, where X is the sum of the percentage of observations removed from both the upper and lower bounds. Thus, the index of the mean is an indication of the percentage of the entries removed on both the sides. For example, if you were to truncate a sample with 8 entries by 12.5%, you would discard the first and the last entry in the sample when calculating the truncated mean.

A trimmed mean for a sample cannot be accurately done. The best is to calculate the nearest two trimmed means, and interpolate (usually linearly).
For example, if you need to calculate the 15% trimmed mean of a sample containing 10 entries, you would calculate the 10% trimmed mean (removing 1 entry on either side of the sample), the 20% trimmed mean (removing 2 entries on either side), and interpolating to determine the 15% trimmed mean.

Let us discuss an example on truncated mean.
For example, a figure skating competition produces the following scores: 6.0, 8.1, 8.3, 9.1, 9.9.
A mean trimmed 40% would equal 8.5 ((8.1+8.3+9.1)/3), which is larger than the arithmetic mean of 8.28. To trim the mean by 40%, we remove the lowest 20% and highest 20% of values i.e. eliminating the scores of 6.0 and 9.1.
As shown by this example, trimming the mean can reduce the effects of outlier bias in a sample.

Let us know the advantage of the truncated mean.
Truncated mean is a useful estimator because it is less sensitive to outliers than the mean. However, it will give a reasonable estimate of central tendency or mean for many statistical models. In this regard, it is referred to as a robust estimator.

Let us know the drawback of the truncated mean.
The truncated mean uses more information from the distribution or sample than the median. However, unless the underlying distribution is symmetric, the truncated mean of a sample is unlikely to produce an unbiased estimator for either the mean or the median.

5. Mid-Range

Now, let us explain the Mid-Range.
Midrange is a basic statistical analysis tool. The midrange determines the number that is directly between the highest and lowest number of your data set.

There is no way to calculate the highest or lowest numbers in your data set. To do this, you need to organize your data in order from highest to lowest or lowest to highest. This helps to reduce the chances of choosing the wrong numbers in the midrange formula.

The midrange of a set of statistical data values is the arithmetic mean of the maximum and minimum values in a data set.
Midrange = (Maximum data entry + Minimum data entry) /2

The mid-range can also be expressed using order statistics. Suppose there is a given set of values X = (X1, X2 …Xn). We take order statistics Y1 to Yn by arranging the Xi's in increasing order, such that Y1 is the minimum value of X and Yn is the maximum value of X. Then, the midrange is defined as:
Mid-range = (Y1 +Yn) / 2

Let us know the disadvantage of the mid-range.
The mid-range is highly sensitive to outliers and ignores all data points by just considering the extreme two data points, which are very sensitive to outliers. Hence, it is a non-robust statistic and it is very rarely used in statistical analysis.
One of the biggest disadvantages of midrange is that one single value can make a big difference in the result.

Let us discuss few examples of mid-range.
Consider the given set of values: 4, 7, 8, 8, 9, 5, 7, 5, 8, 8, 3, 4.
The maximum value of the data is 9 and the minimum value is 2. Then, the Mid-range = (4 + 9)/2 = 6.5. This value gels with the data.

Consider the given set of values: 3, 7, 6, 45, 9, 10, 8, 8. Mid-range = (3 + 45)/2 = 24 which does not gel with the data. This has occurred due to the outlier '45'.

Here's a summary of our learning in this session, where we have understood:

- The concept of weighted mean with respect to arithmetic mean, geometric mean and harmonic mean
- The concept of truncated mean and mid-range