# Arithmatic Mean

By- Dr. Ekta Khare

# Measure of location

- The basic problem of statistics can be stated as follows: Consider a sample of data $x^1, \ldots, x^n$, where $x^1$ corresponds to the first sample point and $x^n$ corresponds to the $n^{th}$ sample point.
- Presuming that the sample is drawn from some population P, what inferences or conclusions can be made about P from the sample?
- Before this question can be answered, the data must be summarized as succinctly as possible; this is because the number of sample points is often large, and it is easy to lose track of the overall picture when looking at individual sample points.
- One type of measure useful for summarizing data defines the center, or middle, of the sample.
- This type of measure is a measure of location/ central tendency.

# The Arithmetic Mean

- **Arithmetic mean** is often referred to as the mean or arithmetic average.
- It is calculated by adding all the numbers in a given data set and then dividing it by the total number of items within that set.
- The arithmetic mean (or mean or sample mean) is usually denoted by $\bar{x}$.
- To calculate the **central tendency** for the given data set, we use different measures like mean, median, mode and so on.
- Among all these measures, the arithmetic mean or mean is considered to be the best measure, because it includes all the values of the data set.
- If any value changes in the data set, this will affect the mean value, but it will not be in the case of median or mode.

# Arithmetic Mean Formula

- The general formula to find the arithmetic mean of a given data is:

  Mean ($\bar{x}$) = Sum of all observations / Number of observations

- It is denoted by $\bar{x}$, (read as x bar).
- Data can be presented in different forms.
- For example, when we have raw data like the marks of a student in five subjects, we add the marks obtained in the five subjects and divide the sum by 5, since there are 5 subjects in total.
- Now consider a case where we have huge data like the heights of 40 students in a class or the number of people visiting an amusement park across each of the seven days of a week.
- Will it be convenient to find the arithmetic mean with the above method? The answer is a big NO! So, how can we find the mean?

# … Arithmetic Mean Formula

- We arrange the data in a form that is meaningful and easy to comprehend.
- Let's understand how to compute the arithmetic average in such cases.
- The arithmetic mean for ungrouped and grouped data can be calculated by the general formula to find the arithmetic mean:

Ungrouped Data: $$\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{n}$$

Grouped Data: $$\bar{x} = \frac{\Sigma f x}{n}$$

Where: $f$ = frequency in each class

$x$ = midpoint of each class

$n$ = total number of scores

## Calculating Arithmetic Mean for Ungrouped Data

Here the arithmetic mean is calculated using the formula:

Mean $\bar{x}$ = Sum of all observations / Number of observations

**Example:** Compute the arithmetic mean of the first 6 odd, natural numbers.

Solution: The first 6 odd, natural numbers: 1, 3, 5, 7, 9, 11

$\bar{x} = (1+3+5+7+9+11) / 6 = 36/6 = 6.$

Thus, the arithmetic mean is 6.

# Calculating Arithmetic Mean for Grouped Data

- There are three methods (Direct method, Short-cut method, and Step-deviation method) to calculate the arithmetic mean for grouped data.
- The choice of the method to be used depends on the numerical value of xi and fi. xi is the sum of all data inputs and fi is the sum of their frequencies.
- ∑ (sigma) the symbol represents summation. If xi and fi are sufficiently small, the direct method will work.
- But, if they are numerically large, we use the assumed arithmetic mean method or step-deviation method.

# Direct Method for Finding the Arithmetic Mean

Let $x_1$, $x_2$, $x_3$ ……$x_{\text{not def}}$ be the observations with the frequency $f_1$, $f_2$, $f_3$ ……$f_{\text{not def}}$.

Then, mean is calculated using the formula:

$\bar{x} = (x_1 f_1 + x_2 f_2 + \ldots\ldots + x_{\text{not def}} f_{\text{not def}}) / \sum f_i$

Here, $f_1 + f_2 + \ldots . f_{\text{not def}} = \sum f_i$ indicates the sum of all frequencies.

**Example I (discrete grouped data):** Find the mean of the following distribution:

| **x** | 10 | 30 | 50 | 70 | 89 |
|---|---|---|---|---|---|
| **f** | 7 | 8 | 10 | 15 | 10 |

**Solution:**

| $x_i$ | $f_i$ | $x_i f_i$ |
|---|---|---|
| 10 | 7 | $10 \times 7 = 70$ |
| 30 | 8 | $30 \times 8 = 240$ |
| 50 | 10 | $50 \times 10 = 500$ |
| 70 | 15 | $70 \times 15 = 1050$ |
| 89 | 10 | $89 \times 10 = 890$ |
| Total | $\sum f_i = 50$ | $\sum x_i f_i = 2750$ |

Add up all the $(x_i f_i)$ values to obtain $\sum x_i f_i$. Add up all the $f_i$ values to get $\sum f_i$

Now, use the mean formula.

$\bar{x} = \sum x_i f_i \,/\, \sum f_i = 2750/50 = 55$

Mean = 55. The above problem is an example of discrete grouped data.

# Data in continuous class intervals

Let's now consider an example where the data is present in the form of **continuous class intervals.**

**Example II (continuous class intervals):** Let's try finding the mean of the following distribution:

| Class - Interval | 15-25 | 25-35 | 35-45 | 45-55 | 55-65 | 65-75 | 75-85 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Frequency | 6 | 11 | 7 | 4 | 4 | 2 | 1 |

**Solution:**

When the data is presented in the form of class intervals, the mid-point of each class (also called class mark) is considered for calculating the mean.

The formula for mean remains the same as discussed above.

Note:

Class Mark = (Upper limit + Lower limit) / 2

| **Class-Interval** | **Class Mark ($x_i$)** | **Frequency ($f_i$)** | **$x_if_i$** |
|---|---|---|---|
| 15-25 | 20 | 6 | 120 |
| 25-35 | 30 | 11 | 330 |
| 35-45 | 40 | 7 | 280 |
| 45-55 | 50 | 4 | 200 |
| 55-65 | 60 | 4 | 240 |
| 65-75 | 70 | 2 | 140 |
| 75-85 | 80 | 1 | 80 |
| | Total | 35 | 1390 |

$\bar{x} = \sum x_if_i / \sum f_i = 1390/35 = 39.71.$

## Short-cut Method for Finding the Arithmetic Mean

The short-cut method is called as **assumed mean method** or **change of origin method**. The following steps describe this method.

**Step1:** Calculate the class marks (mid-point) of each class ($x_i$).

**Step2:** Let **A** denote the assumed mean of the data.

**Step3:** Find deviation (di) = $x_i - A$

**Step4:** Use the formula:

$\bar{x} = A + (\sum f_i d_i / \sum f_i)$

**Example:** Let's understand this with the help of the following example.

Calculate the mean of the following using the short-cut method.

| **Class - Intervals** | 45-50 | 50-55 | 55-60 | 60-65 | 65-70 | 70-75 | 75-80 |
|---|---|---|---|---|---|---|---|
| **Frequency** | 5 | 8 | 30 | 25 | 14 | 12 | 6 |

**Solution:** Let us make the calculation table. Let the assumed mean be **A** = 62.5

Note: A is chosen from the $x_i$ values. Usually, the value which is around the middle is taken.

| **Class-Interval** | Classmark / Mid-points ($x_i$) | **$f_i$** | **$d_i = (x_i - A)$** | **$f_i d_i$** |
|---|---|---|---|---|
| 45-50 | 47.5 | 5 | 47.5-62.5 =-15 | -75 |
| 50-55 | 52.5 | 8 | 52.5-62.5 =-10 | -80 |
| 55-60 | 57.5 | 30 | 57.5-62.5 =-5 | -150 |
| 60-65 | 62.5 | 25 | 62.5-62.5 =0 | 0 |
| 65-70 | 67.5 | 14 | 67.5-62.5 =5 | 70 |
| 70-75 | 72.5 | 12 | 72.5-62.5 =10 | 120 |
| 75-80 | 77.5 | 6 | 77.5-62.5 =15 | 90 |
| | | $\sum f_i$=100 | | $\sum f_i d_i$= -25 |

Now we use the formula,

$\bar{x} = A + (\sum f_i d_i / \sum f_i) = 62.5 + (-25/100) = 62.5 - 0.25 = 62.25$

$\therefore$ Mean = 62.25

## Step Deviation Method for Finding the Arithmetic Mean

This is also called the **change of origin or scale method.** The following steps describe this method:

Step 1: Calculate the class marks of each class ($x_i$).

Step 2: Let **A** denote the assumed mean of the data.

Step 3: Find $u_i = (x_i - A)/h$, where h is the class size.

Step 4: Use the formula:

$\bar{x} = A + h \times (\sum f_i u_i / \sum f_i)$

**Example:** Consider the following example to understand this method. Find the arithmetic mean of the following using the step-deviation method.

| **Class Intervals** | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | Total |
|---|---|---|---|---|---|---|---|---|
| **Frequency** | 4 | 4 | 7 | 10 | 12 | 8 | 5 | 50 |

**Solution:** To find the mean, we first have to find the class marks and decide A (assumed mean). Let A = 35 Here h (class width) = 10

| C.I. | $x_i$ | $f_i$ | Ui = (xi-A)/h | $f_iu_i$ |
|---|---|---|---|---|
| 0-10 | 5 | 4 | -3 | 4 x (-3)=-12 |
| 10-20 | 15 | 4 | -2 | 4 x (-2)=-8 |
| 20-30 | 25 | 7 | -1 | 7 x (-1)=-7 |
| 30-40 | 35 | 10 | 0 | 10 x 0= 0 |
| 40-50 | 45 | 12 | 1 | 12 x 1=12 |
| 50-60 | 55 | 8 | 2 | 8 x 2=16 |
| 60-70 | 65 | 5 | 3 | 5 x 3=15 |
| Total | | ∑fi=50 | | ∑fiui=16 |

Using mean formula:

$\bar{x} = A + h \times (\sum f_iu_i/\sum f_i) = 35 + (16/50) \times 10 = 35 + 3.2 = 38.2$

Mean = 38.

# Advantages of Arithmetic Mean

- As the formula to find the arithmetic mean is rigid, the result doesn't change. Unlike the median, it doesn't get affected by the position of the value in the data set.
- It takes into consideration each value of the data set.
- Finding an arithmetic mean is quite simple; even a common man having very little finance and math skills can calculate it.
- It's also a useful measure of central tendency, as it tends to provide useful results, even with large groupings of numbers.
- It can be further subjected to many algebraic treatments, unlike mode and median. For example, the mean of two or more series can be obtained from the mean of the individual series.
- The arithmetic mean is widely used in geometry as well. For example, the coordinates of the "centroid" of a triangle (or any other figure bounded by line segments) are the arithmetic mean of the coordinates of the vertices.

# Disadvantages of Arithmetic Mean

- The strongest drawback of arithmetic mean is that it is affected by extreme values in the data set. To understand this, consider the following example. It's Ryma's birthday and she is planning to give return gifts to all who attend her party. She wants to consider the mean age to decide what gift she could give everyone. The ages (in years) of the invitees are as follows: 2, 3, 7, 7, 9, 10, 13, 13, 14, 14 Here, n = 10. Sum of the ages = 2+3+7+7+9+10+13+13+14+14 = 92. Thus, mean = 92/10 = 9.2 In this case, we can say that a gift that is desirable to a kid who is 9 years old may not be suitable for a child aged 2 or 14.

- In a distribution containing open-end classes, the value of the mean cannot be computed without making assumptions regarding the size of the class.

| Class Interval | Frequency |
|---|---|
| Less than 15 | 20 |
| 15-25 | 12 |
| 25-35 | 3 |
| 35-45 | 12 |
| More than 45 | 6 |

We know that to find the arithmetic mean of grouped data, we need the mid-point of every class. As evident from the table, there are two cases (less than 15 and 45 or more) where it is not possible to find the mid-point and hence, arithmetic mean can't be calculated for such cases.

- It's practically impossible to locate the arithmetic mean by inspection or graphically.
- It cannot be used for qualitative types of data such as honesty, favorite milkshake flavor, most popular product, etc.
- We can't find the arithmetic mean if a single observation is missing or lost.