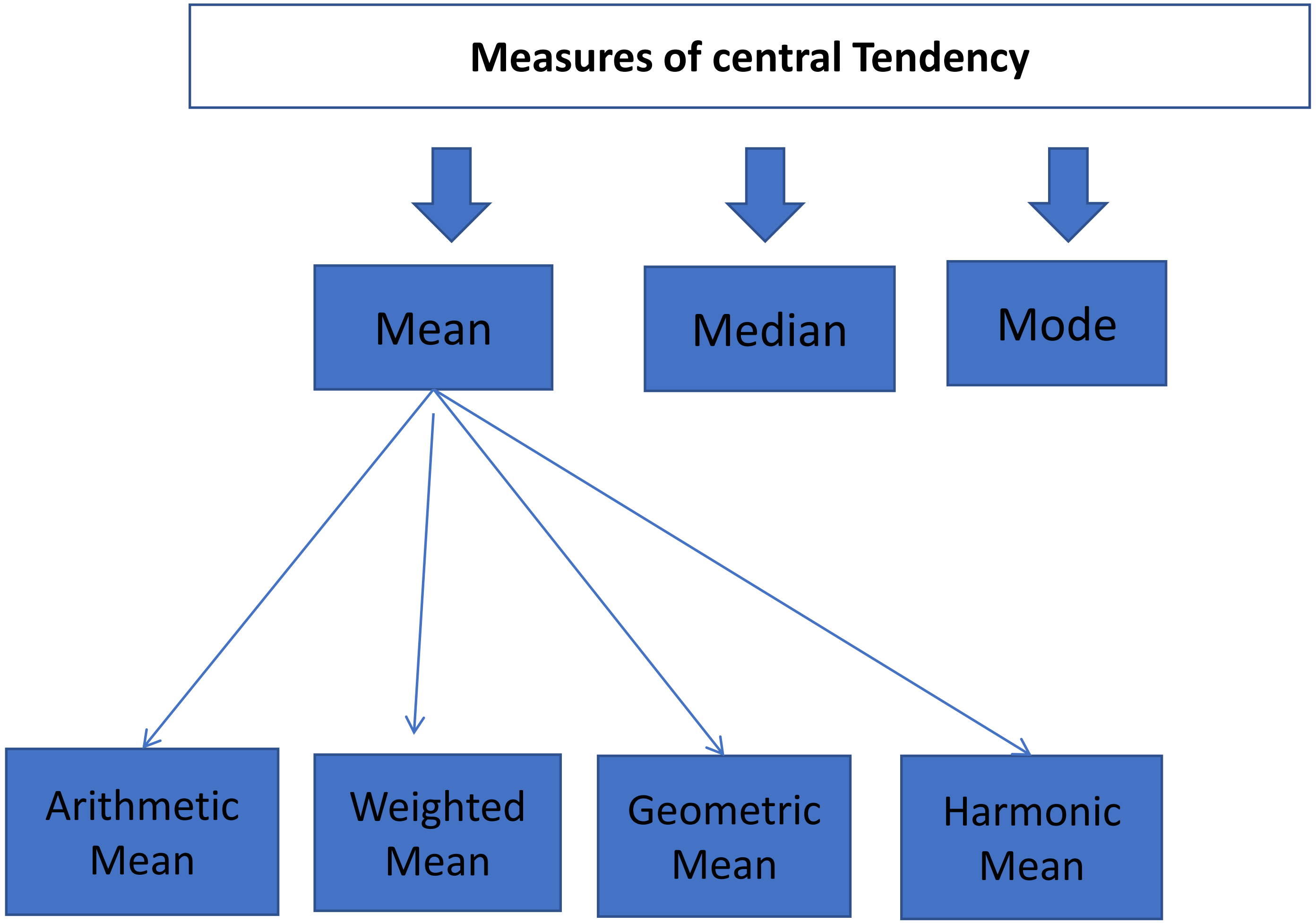
Measures of Central Tendency

Shahrear Khan Rasel
Senior Lecturer
Dept. of GED
Daffodil International University

Measures of central tendency

- A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data.
- As such, measures of central tendency are sometimes called measures of central location.
- They are also classed as summary statistics.
- The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others.
- In the following sections, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.
- The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the median and the mode.


Measures of central Tendency
Mean
Median
Mode
Arithmetic Mean
Weighted Mean
Geometric Mean
Harmonic Mean

Mean (Arithmetic) $\overline{x}$

- The mean (or average) is the most popular and well known measure of central tendency.
- It can be used with both discrete and continuous data, although its use is most often with continuous data.
- The mean is equal to the sum of all the values in the data set divided by the number of values in the data set.

Formulas

- **Ungrouped Data:** Let x be a random variable and the x values are given by $\{x_1, x_2, \ldots, x_n\}$. Now, the arithmetic mean for n ungrouped observations is given by:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

- **Grouped Data:** Let x be a random variable and the mid values of x values are given by $\{x_1, x_2, \ldots, x_n\}$ and the respective frequencies are $\{f_1, f_2, \ldots, f_n\}$. Now, the arithmetic mean for n grouped observations is given by:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{m} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{m} \; ; \mathrm{m} = \sum_{i=1}^{n} f_i$$

Merits & Demerits

- **Merits of Arithmetic Mean:**

➢ It is easy to understand and also easy to calculate.

➢ It takes all the observations into account.

- **Demerits of Arithmetic Mean:**

➢ It is affected by extreme values.

➢ It cannot be computed from a distribution that has the open-end class interval.

Types of Mean

- **Weighted Arithmetic Mean:** The weighted mean is a special case of the arithmetic mean. It occurs when there are several observations of the same value.
- The weighted mean of the positive real numbers $\{x_1, x_2, \dots, x_n\}$ with their weight $\{w_1, w_2, \dots, w_n\}$ is defined by
- $\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{n} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{n}; n = \sum_{i=1}^{n} w_i$
- **Geometric Mean (GM):** The geometric mean is a type of average, usually used for growth rates, like population growth or interest rates. While the arithmetic mean **adds** items, the geometric mean **multiplies** items. Also, you can only get the geometric mean for positive numbers.
- [***Ungrouped data***] The geometric mean of the n observations $\{x_1, x_2, \dots, x_n\}$ is defined by
- GM= $(x_1 \times x_2 \times \cdots \times x_n)^{1/n}$
- [***Grouped data***] The geometric mean of the n observations $\{x_1, x_2, \dots, x_n\}$ with the respective frequencies are $\{f_1, f_2, \dots, f_n\}$ is defined by
- GM= $\left(x_1^{f_1} \times x_2^{f_n} \times \cdots \times x_n^{f_n}\right)^{1/m}; m = \sum_{i=1}^{n} f_i$

- **Harmonic Mean (HM):** The harmonic mean (sometimes called the sub-contrary mean) is a very specific type of average and in particular one of the Pythagorean means. It's generally used when dealing with averages of units, like speed or other rates and ratios.

- **[Ungrouped Data]** The harmonic mean of the n numbers $\{x_1, x_2, \dots, x_n\}$ is defined by
- $HM = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\dots+\frac{1}{x_n}}$
- **[Grouped Data]** The harmonic mean of the n observations $\{x_1, x_2, \dots, x_n\}$ with the respective frequencies are $\{f_1, f_2, \dots, f_n\}$ is defined by
- $HM = \frac{m}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\dots+\frac{f_n}{x_n}}$; $m = \sum_{i=1}^{n} f_i$

Example 1: The weight of the row materials for producing a drug by 10 companies is as follows(in ml):

63 61 65 62 61 64 60 66 68 61

Find out the (i) mean/average/arithmetic (ii) geometric and (iii) harmonic mean of the given data.

Solution(i): Here, n=10, Let the mean be $\bar{x}$

$$\therefore \bar{x} = \frac{63+61+65+62+61+64+60+66+68+61}{10} = 63.1ml \text{ (ans.)}$$

Solution(ii): Here, n=10, Let the mean be $\bar{x}$

$$\therefore \bar{x} = (63 \times 61 \times 65 \times 62 \times 61 \times 64 \times 60 \times 66 \times 68 \times 61)^{1/10} = 63.052\ ml \text{ (ans.)}$$

Solution(iii): Here, n=10,

$$\therefore HM = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}} = \frac{10}{\frac{1}{63}+\frac{1}{61}+\frac{1}{65}+\frac{1}{62}+\frac{1}{61}+\frac{1}{64}+\frac{1}{60}+\frac{1}{66}+\frac{1}{68}+\frac{1}{61}} = 62.893\ ml \text{ (ans.)}$$

Example 2: Let the number of failure of a test between the ages is as follows. Find the average failure rate.

Class Interval	10 - 15	15 - 20	20 - 25	25 - 30
Frequency	12	14	20	8

Solution: we first construct the following table:

Class Interval	Frequency(f_i)	Mid value(x_i)	$f_i x_i$
10-15	12	12.5	150
15-20	14	17.5	245
20-25	20	22.5	450
25-30	8	27.5	220
Total	$m = \sum f_i = 54$		$\sum f_i x_i = 1065$

Here, m=54, Let the mean be $\bar{x}$

$$\therefore \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{m} = \frac{1065}{54} = 19.73 \text{ (ans.)}$$

Example 3: Let the number of failure of a test between the ages is as follows. Find the (i) geometric and (ii) harmonic average failure rate.

Class Interval	10 - 15	15 - 20	20 - 25	25 - 30
Frequency	12	14	20	8

Solution(i) : we first construct the following table:

Class Interval	Frequency(f_i)	Mid value(x_i)
10-15	12	12.5
15-20	14	17.5
20-25	20	22.5
25-30	8	27.5
Total	$m = \sum f_i = 54$	

Here, m=54, Let the mean be $\bar{x}$

$$\therefore \bar{x} = \left(x_1^{f_1} \times x_2^{f_n} \times \cdots \times x_n^{f_n}\right)^{1/m} = [(12.5)^{12} \times (17.5)^{14} \times (22.5)^{20} \times (27.5)^{8}]^{1/54}$$

$= 19.057$(ans.)

Solution(ii) :

Here, m=54,

$\therefore$ HM= $\frac{m}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\cdots+\frac{f_n}{x_n}} = \frac{54}{\frac{12.5}{12}+\frac{17.5}{14}+\frac{22.5}{20}+\frac{27.5}{8}} = 19.72\ (ans.)$

Example 4: A restaurant sold medium, large and big size soft drinks for \$.90, \$1.25, and \$1.50 respectively. Of the last 10 drinks sold, 3 were medium, 4 were large and 3 were big sized. Find the mean price of the last 10 drinks sold.

- **Solution:** This is a problem of weighted average.

Here, $x_1 = 3,\ x_2 = 4,\ x_3 = 3\ and\ w_1 = 0.90,\ w_2 = 1.25,\ w_3 = 1.50$

We can apply the weighted mean,

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3}{w_1 + w_2 + w_3}$$

$$= \frac{3(0.90) + 4(1.25) + 3(1.50)}{10}$$

$$= 1.22$$

So, the mean price is \$1.22.

Median

- The middle value that separates the higher half from the lower half of the data set.
- The median and the mode are the only measures of central tendency that can be used for ordinal data, in which values are ranked relative to each other but are not measured absolutely.
- It’s used for many real-life situations, like Bankruptcy law, where you can only claim bankruptcy if you are below the median income in their state.

Median For Ungrouped Data:

The Median is defined as the middle most observation when the observations are arranged in order of magnitude (in ascending or descending order).

- when 'n' is odd, the middle-most observation that is $\left(\frac{n+1}{2}\right)^{th}$ observation will be the median in the series. That is, Median = $\left(\frac{n+1}{2}\right)^{th}$ observation, when 'n' is odd.
- when 'n' is even, the median will be the arithmetic mean of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations in the series. That is, Median $= \left[\frac{\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n}{2}+1\right)^{th} observation}{2}\right]$ when 'n' is even.

Median For Grouped Data:

For grouped frequency distribution the Median is given by $Me = L + \frac{c}{f}\left(\frac{n}{2} - F\right)$

where L = The lower limit of the median class

F = Cumulative frequency of the class just preceding the median class

f = Frequency of the median class

c = Length of the median class

Example 5: What is the median of the following observations?
(i) 12, 15, 14, 78, 56
(ii) 63, 61, 65, 62, 61, 64, 60, 66, 68, 61

Solution (i): At first, we sort the data in ascending order, 12, 14, 15, 56, 78

Here, the number of observations, $n = 5$ which is odd so the median is

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} = \left(\frac{5+1}{2}\right)^{th} = 3^{rd} \text{ observation} = 15$$

So the median is 15.

Solution:

To calculate the median, the data have to be arranged either in ascending or descending order. Here, the data has been arranged in ascending order.

60	61	61	61	**62**	**63**	64	65	66	68

We know that, when n is even, the median will be the arithmetic mean of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations in the series. In this problem we have, $n=10$, so the $\left(\frac{n}{2}\right)^{th}$ observation is $\left(\frac{10}{2}\right)^{th} = 5^{th}$ observation and $\left(\frac{n+1}{2}\right)^{th}$ observation is $\left(\frac{10}{2}+1\right)^{th} = 6^{th}$ observation.

Median = (5^{th} observation + 6^{th} observation)/2 = (62+63)/2 = 62.5

Example 6: The following table gives the weekly wages in Taka: Find out the median of the above problem.

Weekly wages (Tk.)	50-53	53-56	56-59	59-62	62-65	65-68	68-71	71-74	74-77
No. of persons	3	8	14	30	36	28	16	10	5

Solution: Table for calculation of Median

Weekly wages (Tk.)	No. of persons or Frequency (*f*)	Cumulative Frequency (*F*)
50-53	3	3
53-56	8	11
56-59	14	25
59-62	30	55
62-65	36	91
65-68	28	119
68-71	16	135
71-74	10	145
74-77	5	150
Total	n = 150	

Here, $n = 150$. Therefore $n/2 = 75$. Looking at the cumulative range column in the table, we find that $n/2$ falls in the range 62 – 65. So the median class is 62 – 65.

We know median, $Me = L + \frac{c}{f}\left(\frac{n}{2} - F\right)$

Where, L = the lower limit of the median class = 62

c = the class width of the median class = 3

n = the total number of observations = 150

F = the cumulative frequency preceding to the median class = 55

f = the frequency of the median class = 36

So, median, $Me = 62 + \frac{3}{36} \times \left(\frac{150}{2} - 55\right) = 63.67$

Mode

- The most frequent value in the data set.
- This is the only central tendency measure that can be used with nominal data, which have purely qualitative category assignments.
- **Mode For Ungrouped Data:** For raw data, the mode is the value corresponding to the maximum frequency. Sometimes there may be no single mode if no one value appears more than any other. There may also be two modes (bimodal), three modes (trimodal), or more than three modes (multi-modal).

- **Example 7:** Find the mode of the series 2, 5, 2, 5, 7, 8, 5.

- **Solution:** Here the observation 5 occurs maximum 3 times. So, the mode is Mo = 5.

Mode For Grouped Data:

For grouped data, the modal class is the class with the largest frequency. After identifying modal class, the mode is evaluated by using the following formula

$$Mode,\ Mo = l + \frac{d_1}{d_1 + d_2} \times C \quad \text{or} \quad Mode,\ Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

Where l is the lower limit of the modal class for which the frequency is maximum

C is the class width of the modal class

d_1 is the difference between the frequency of the modal class and pre modal class

d_2 is the difference between the frequency of the modal class and post modal class.

Example 8: Calculate the mode for the following frequency distribution:

Class	9.5-14.5	14.5-19.5	19.5-24.5	24.5-29.5	29.5-34.5
Frequency(f)	7	12	5	6	8

Solution: Since the maximum frequency (12) lies in class 14.5 – 19.5. So the modal class is 14.5 – 19.5

We know, $Mode,\ Mo = l + \frac{d_1}{d_1 + d_2} \times C$

Here, l = the lower limit of the modal class =14.5

C = the class width of the modal class = 5

d_1 = the difference between the frequency of the modal class and pre modal class = 12 – 7 =5

d_2 = the difference between the frequency of the modal class and post modal class = 12 – 5 = 7

So, $Mode,\ Mo = l + \frac{d_1}{d_1 + d_2} \times C = 14.5 + \frac{5}{5+7} 5 = 16.58$

- **Characteristics to be a good measure of central tendency:**

1) Should be easy to compute and easily understandable.
2) Should be amenable to mathematical calculation.
3) Should be based on all observation.
4) Should be unaffected by the presence of extreme values
5) Should not have sampling variability.

- **Compare Mean, Median, and Mode:**

1) Mean, median and mode are easy to understand and easy to calculate.
2) Mean is based upon all observations but median and mode are not.
3) Mean is amenable to mathematical calculation but median and mode are not.
4) Mean cannot be calculated from the distribution with open class but median and mode can be calculated from the distribution with classes.
5) Mean is affected very much by extreme values but median and mode are not affected by the extreme values.

Please use the following summary table to know what the best measure of central tendency is with respect to the different types of variable.

Type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

Theoretical Questions:

1. Write down the short note about the following terms:
    a) Central tendency,
    b) Arithmetic mean,
    c) Weighted arithmetic mean,
    d) Geometric mean,
    e) Harmonic mean,
    f) Median,
    g) Mode
    h) Quartiles
    i) Deciles
    j) Percentiles

2. Classify the different types of measures of central tendency.

Mathematical Problems:

1. In a class of 30 students, 4 students were present in the class for 25 days in the month of March, 15 were present for 22 days, 5 were present for 18 days and the remaining 6 were present for 10 days only. Calculate the number of days a student was present on the average during the month of March. (Answer: 19.3 or 20 days)
2. The weights of 7 mothers in kg were recorded as follows:

   47, 52, 55, 60, 48, 42, 61.

   Find the arithmetic mean, geometric mean, harmonic mean, median, mode, quartiles (Q_2), deciles (D_6), and percentiles (P_{60}) of the given problem.

3. The number of students presents in a class in 20 working days:

20	20	26	30	22	15	17	34	23	35	23	36	28
34	31	19	38	32	37	24						

Find the arithmetic mean, geometric mean, harmonic mean, median, mode, quartiles (Q_2), deciles (D_6), and percentiles (P_{60}) of the given problem.

4. The thickness of the printed circuit board is very important characteristics. A sample of 10 boards had the following thickness (in thousands of an inch):

63, 61, 65, 62, 61, 64, 60, 66, 68, 61

Find the arithmetic mean, geometric mean, harmonic mean, median, mode, quartiles (Q_3), deciles (D_7), and percentiles (P_{84}) of the given problem.

5. For the following frequency distribution find out the arithmetic mean, geometric mean, harmonic mean, median, mode, quartiles (Q_2), deciles (D_6), and percentiles (P_{60}) of the given problem.

Class	9.5-14.5	14.5-19.5	19.5-24.5	24.5-29.5	29.5-34.5
Frequency	7	12	5	6	8

14. The following table gives the weekly wages in taka.

Weekly wages (Tk.)	No. of persons
50-53	3
53-56	8
56-59	14
59-62	30
62-65	36
65-68	28
68-71	16
71-74	10
74-77	5

find out the arithmetic mean, geometric mean, harmonic mean, median, mode, quartiles (Q_2), deciles (D_6), and percentiles (P_{60}) of the given problem.

15. The wage data of Beximco pharmaceutical company are given below:

Weakly wage in Tk.	**Frequency**
48.5 – 53.5	2
53.5 – 58.5	2
58.5 – 63.5	3
63.5 – 68.5	5
68.5 – 73.5	5
73.5 – 78.5	5
78.5 – 83.5	5
83.5 – 88.5	7
88.5 – 93.5	10
93.5 – 98.5	6

Find the arithmetic mean, geometric mean, harmonic mean, median, mode, quartiles (Q_3), deciles (D_4), and percentiles (P_{46}) of the given problem.