Measure Of Central Tendency

There are four characteristics of a data set or a frequency distribution. They are

- Location or central tendency
- Dispersion
- Skewness
- Kurtosis

Some Definitions

Simpson and Kafka defined it as " A measure of central tendency is a typical value around which other figures gather"

Waugh has expressed "An average stand for the whole group of which it forms a part yet represents the whole".

In layman's term, a measure of central tendency is an AVERAGE. It is a single number of value which can be considered typical in a set of data as a whole.

Characteristics of a Good Measure of Central Tendency

- It should be easy to understand
- It should be easy to compute
- It should be rigidly defined. This says that, the series of whose average is calculated should have only one interpretation. One interpretation will avoid personal prejudice or bias.
- It should be based on all observations. In other wards, the value should lie between the upper and lower limit of the data.
- It should be capable of further algebraic treatment. In other wards, an ideal average is one which can be used for further statistical calculations.
- It should have sampling stability
- It should not be affected by the presence of extreme values

Mean

- The MEAN of a set of values or measurements is the total or sum of all the measurements divided by the number of measurements in the set.

- The mean is the most popular and widely used.
- It is sometimes called the arithmetic mean.
- It is the best measure of central tendency

Properties of Mean

1. Mean can be calculated for any set of numerical data, so it always exists.
2. A set of numerical data has one and only one mean.
3. Mean is the most reliable measure of central tendency since it takes into account every item in the set of data.
4. It is used only if the data are interval or ratio.

Demerits or Limitations

- It is greatly affected by extreme or deviant values (*outliers*)
- It can’t be calculated for qualitative data
- It can’t be found graphically
- It can’t be computed in case of open-ended class interval of a frequency distribution
- It is not a good measure of central tendency in case of highly shewed distribution

Mean for Ungrouped data

- If we get the mean of the sample, we call it the sample mean and it is denoted by (read "x bar").

$$\text{Sample mean} = \frac{\text{Sum of all the values in the sample}}{\text{Number of values in the sample}}$$

- If we compute the mean of the population, we call it the parametric or population mean, denoted by μ (read "mu").

$$\text{Population mean} = \frac{\text{Sum of all the values in the population}}{\text{Number of values in the population}}$$

Median

- The **median** is the middle most value of a set of observations when the values are arranged in order of magnitude.
- Connor has defined as " The median is that value of the variable which divides the group into two equal parts, one part comprising of all values greater, and the other, all values less than median"
- For Ungrouped data median is calculated as:

$$Median(M) = Size\ of \left(\frac{N+1}{2}\right) th\ item$$

- For Grouped Data:

$$Median(M) = L1 + \frac{\frac{N}{2} - cf}{f} \times i$$

Advantages of Median:

- Median can be calculated in all distributions.
- Median can be understood even by common people.
- Median can be ascertained even with the extreme items.
- It can be located graphically
- It is most useful dealing with qualitative data

Disadvantages of Median

- It is not based on all the values.
- It is not capable of further mathematical treatment.
- It is affected fluctuation of sampling.
- In case of even no. of values it may not the value from the data.

Example of median For Grouped Data

Marks	No. of students	Less than CF
0 – 4	2	2
5 – 9	8	10
10 – 14	14	24
15 – 19	17	41
20 – 24	9	50
Total	50	

As median $= L + \frac{h}{f}\left(\frac{n}{2} - c\right)$

$n=50;\ h=5\ ;L=14.5$

$f=17$ *and* $c=24$

so

$$Median = 14.5 + \frac{5}{17}\left(\frac{50}{2} - 24\right)$$

$$= 14.94$$

Mode

- The mode, denoted Mo, is the value which occurs most frequently in a set of measurements or values. In other words, it is the most popular value in a given set.

- Croxton and Cowden : defined it as “the mode of a distribution is the value at the point armed with the item tend to most heavily concentrated. It may be regarded as the most typical of a series of value”

Properties of mode

- It is used when you want to find the value which occurs most often.
- It is a quick approximation of the average.
- It is an inspection average.
- It is the most unreliable among the three measures of central tendency because its value is undefined in some observations.

Advantages of Mode

- Mode is readily comprehensible and easily calculated
- It is the best representative of data
- It is not at all affected by extreme value.
- The value of mode can also be determined graphically.
- It is usually an actual value of an important part of the series.

Disadvantages of Mode

- It is not based on all observations.
- It is not capable of further mathematical manipulation.
- Mode is affected to a great extent by sampling fluctuations.
- Choice of grouping has great influence on the value of mode.

Geometric Mean

In business and economic problems, very often we face the questions pertaining to percentage, rate of change of over time. In that case , neither the mean, median nor mode is the appropriate measure of location. The geometric mean is a useful measure of the average rate of change of a variable over time

Geometric Mean

- Geometric mean is a kind of average of a set of numbers that is different from the arithmetic average.
- The geometric mean is well defined only for sets of positive real numbers. This is calculated by multiplying all the numbers (call the number of numbers n), and taking the nth root of the total.
- A common example of where the geometric mean is the correct choice is when averaging growth rates.
- The geometric mean is NOT the arithmetic mean and it is NOT a simple average.

Definition

The geometric mean of n non-zero and positive values is the nth root of their product.

Suppose $x_{1,}\, x_2, \ldots \ldots x_n$ are n non-zero observations of a data set. Then

$$G.M. = \sqrt[n]{x_{1.}x_2 \ldots \ldots x_n}$$

Harmonic Mean

Harmonic mean of non-zero values is the reciprocal of the arithmetic mean of the reciprocal of the individuals.

Suppose $x_{1,}\, x_2, \dots \dots x_n$ are n non-zero observations of a data set.

- ❑ For ungrouped data

$$H.M. = \frac{n}{\sum \frac{1}{x}}$$

- ❑ For grouped data

$$H.M. = \frac{n}{\sum \frac{f}{x}}$$

Arithmetic Mean Calculated Methods for grouped data:

- **Direct Method :**

$$\overline{X} = \frac{\sum fm}{N}$$

- **Short cut method :**

$$\overline{X} = A + \frac{\sum fd}{N}$$

- **Step deviation Method :**

$$\overline{X} = A + \frac{\sum fd}{N} \times i$$

Example-1

A sample of five executives received the following bonus last year ($000):

14.0, 15.0, 17.0, 16.0, 15.0

Compute arithmetic mean

Solution:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{14.0 + \ldots + 15.0}{5} = \frac{77}{5} = 15.4$$

Example 2: In a crash test, 11 cars were tested to determine what impact speed was required to obtain minimal bumper damage. Find the mode of the speeds given in miles per hour below.

24, 15, 18, 20, 18, 22, 24, 26, 18, 26, 24

Solution: Ordering the data from least to greatest, we get:

15, 18, 18, 18, 20, 22, 24, 24, 24, 26, 26

Answer: Since both 18 and 24 occur three times, the modes are 18 and 24 miles per hour.

Example-3

The following frequency distribution refers to the number of hours worked per month of 50 workers of a factory. Compute mean, median, mode

No. of hours worked per month	Frequency
30-55	3
55-80	4
80-105	6
105-130	9
130-155	12
155-180	11
180-205	5

Solution

Class	Frequency f	Mid point x	fx	Cumulative frequency
30-55	3	42.5	127.5	3
55-80	4	67.5	270	7
80-105	6	92.5		13
105-130	9	117.5		22
130-155	12	142.5		34
155-180	11	167.5		45
180-205	5	192.5		50
Total	n =50		$\sum fx = 6525$	

Mean:

$$\bar{x} = \frac{\sum fx}{n} = 130.5$$

Mode:

$$Mo = L + \frac{f_1}{f_1 + f_2} \times i$$

Highest frequency is 12 which lies in the class 130-155 which is the modal class

L=130, $f_1 = 12 - 9 = 3$,

f_2=12-11=1, i=25

Then $Mo = 148.7$

Median

$$Me = L + \frac{\frac{n}{2} - f_c}{f} \times i$$

$\frac{n}{2}$=25th item lies in the class 130-155 which is the median class

L=130, f=12, f_c=22, i=25.

$Me = 136.25$

Theorem-1: For two non-zero quantities

$$A.M \geq G.M. \geq H.M.$$

Theorem-2: For two non-zero quantities

$$G.M. = \sqrt{A.M \ * H.M.}$$

Example: Suppose the geometric and harmonic mean of two positive quantities are $4\sqrt{3}, 6$.Find their arithmetic mean.

Ans.: 8

Weighted Mean

- **Weighted mean** is the mean of a set of values wherein each value or measurement has a different weight or degree of importance. The following is its formula:

$$\bar{x} = \frac{\sum xw}{\sum w}$$

where $\bar{x}$ = mean

x = measurement or value

w = number of measurements

Example Of W.M

Below are Amaya's subjects and the corresponding number of units and grades she got for the previous grading period. Compute her grade point average.

Subject	Units	Grade
Filipino	.9	86
English	1.5	85
Mathematics	1.5	88
Science	1.8	87
Social Studies	.9	86
TLE	1.2	83
MAPEH	1.2	87

$$\bar{X} = \frac{(0.9*86)+(1.5*85)+(1.5*88)+(1.8*87)+(0.9*86)+(1.2*83)+(1.2*87)}{9}$$

$$= 86.1$$

Amaya's average grade is 86.1

Harmonic Mean Example

Calculate the harmonic mean of the numbers: 13.2, 14.2, 14.8, 15.2 and 16.1

Solution:

The harmonic mean is calculated as below:

AS

$$H.M \text{ of } X = \overline{X} = \frac{n}{\sum\left(\frac{1}{x}\right)}$$

$$H.M \text{ of } X = \overline{X} = \frac{5}{0.3417} = 14.63$$

x	0.0658 0.0621
13.2	0.0758
14.2	0.0704
14.8	0.0676
15.2	0.0658
16.1	0.0621
Total	$\sum\frac{1}{x}$ =0.3147

Example: Calculate the harmonic mean for the given below:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
F	2	3	11	20	32	25	7

Solution: Now We'll find H.M as:

Marks	x	f	$\frac{f}{x}$
30-39	34.5	2	0.0580
40-49	44.5	3	0.0674
50-59	54.5	11	0.2018
60-69	64.5	20	0.3101
70-79	74.5	32	0.4295
80-89	84.5	25	0.2959
90-99	94.5	7	0.0741
Total		4.5	0.2959

$$\overline{X} = \frac{\sum f}{\sum\left(\frac{f}{x}\right)} = \frac{100}{1.4368} = 69.60$$

Formulas

- For Grouped data:

$$G=\text{anti}\left(\sum \frac{f \log x}{n}\right)$$

Question 1: Find the geometric mean of the following values:

15, 12, 13, 19, 10

$$n=5, \sum \log x = 5.648$$

As $$G = \text{Anti}\left(\frac{\sum \log x}{n}\right)$$

$$G = Anti\left(\frac{5.648}{5}\right)$$

$$G = Anti(1.129)$$

$$G = 13.48$$

x	*Log x*
15	1.1761
12	1.0792
13	1.1139
19	1.2788
10	1.0000
Total	5.648

Formula:

- fn is the frequency of the modal class
- f1 is the frequency of the class before the modal class in the frequency table
- f2 is the frequency of the class after the modal class in the frequency table
- h is the class interval of the modal class

Find the modal class and the actual mode of the data set below

Number	Frequency
1 - 3	7
4 - 6	6
7 - 9	4
10 - 12	2
13 - 15	2
16 - 18	8
19 - 21	1
22 - 24	2
25 - 27	3
28 - 30	2

Here model class=10-12

L = 10, fn= 9, f1 = 4, f2 = 2, h = 3

$$\text{Mode}=L+\left(\frac{fn-f1}{(fn-f1)+(fn-f2)}\right)*h$$

$$=10+\left(\frac{9-4}{(9-4)+(9-2)}\right)*3$$

$$=10+\left(\frac{5}{5+7}\right)*3$$

$$=10+1.25$$

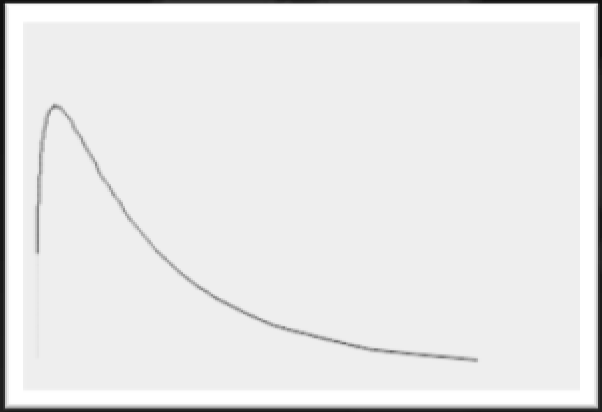
Mode=11.25

Relations Between the Measures of Central Tendency

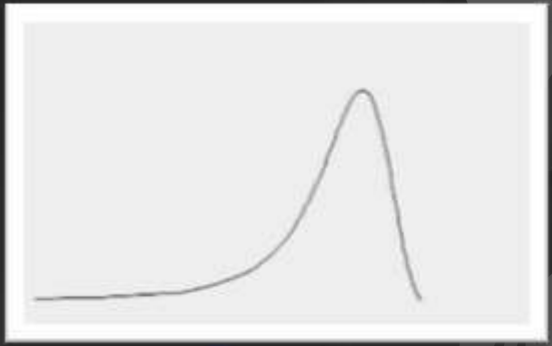
- In symmetrical distributions, the median and mean are equal

For normal distributions, mean = median = mode

- In positively skewed distributions, the mean is greater than the median



- In negatively skewed distributions, the mean is smaller than the median



Conclusion

- A measure of central tendency is a measure that tells us where the middle of a bunch of data lies.
- Mean is the most common measure of central tendency. It is simply the sum of the numbers divided by the number of numbers in a set of data. This is also known as average.
- Median is the number present in the middle when the numbers in a set of data are arranged in ascending or descending order. If the number of numbers in a data set is even, then the median is the mean of the two middle numbers.
- Mode is the value that occurs most frequently in a set of data.

Thank You… ☺ ☺

☺