**Understanding Tendency of the Data**

Requisites for an ideal Measure of Central Tendency.

According to Professor Yule

The following are the characteristics to be satisfied by an ideal measure of central tendency.

1. It should be rigidly defined.
2. It should be readily comprehensible and easy to calculate.
3. It should be based on all the observations.
4. It should be suitable for further mathematical treatment.
5. It should be affected as little as possible by fluctuations of sampling.
6. It should not be affected much by extreme values.

The following are the popular measures of central tendency.

**The Mean**

For a given set of numbers, the most familiar and useful measure of the center is the mean, or arithmetic average of the set. Because we will almost always think of the xi’s as constituting a sample, we will often refer to the arithmetic average as the sample mean and denote it by.

The sample mean of observations is given by

= =

For ex: 9+8+6+4+4+10+12+15= 68 = 8.5

8 8

The numerator of can be written more informally as, where the summation is over all sample observations.

**In case of frequency distribution**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | x1 | x2 | …… | xn |
| f | f1 | f2 | …… | fn |

The arithmetic mean =∑ fi\*x / ∑ fi



For example, the following is the frequency distribution of the number of telephone calls received in 245 successive 1-minute intervals at an exchange. Calculate the mean incoming calls per minute.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. of calls | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Frequency | 14 | 21 | 25 | 43 | 51 | 40 | 39 | 12 |
| fi\*xi | 0 | 21 | 50 | 129 | 204 | 200 | 234 | 84 |



=∑ fi\*x/ ∑ fi = 922/245=3.763

**Calculating mean from grouped data**

To find arithmetic of grouped data we first calculate the midpoints of each class (if require we round up the values)

1. Then we multiply each midpoint by frequency of that observation in that class.
2. Sum all the results.
3. Divide the sum by total no of observations.

= where

= sample mean.

∑ = symbol meaning “the sum of”

=frequency (number of observations) in each class.

=mid point for each class in the sample.

=number of observations in the sample.

The following is the frequency distribution

For example: The income per month of 600 families in a particular region is grouped as follows

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class | 0-49 | 50-99 | 100-149 | 150-199 | 200-249 | 250-299 | 300-349 | 350-399 | 400-449 | 450-499 | Total |
| Frequency | 75 | 126 | 180 | 89 | 50 | 48 | 10 | 12 | 5 | 5 | 600 |

|  |  |  |  |
| --- | --- | --- | --- |
| Class | Midpoint(x) | Frequency(f) | f\*x |
| 0-49 | 25 | 75 | 1875 |
| 50-99 | 75 | 126 | 9450 |
| 100-149 | 125 | 180 | 22500 |
| 150-199 | 175 | 89 | 15575 |
| 200-249 | 225 | 50 | 11250 |
| 250-299 | 275 | 48 | 13200 |
| 300-349 | 325 | 10 | 3250 |
| 350-399 | 375 | 12 | 4500 |
| 400-449 | 425 | 5 | 2125 |
| 450-499 | 575 | 5 | 2875 |
| Total |  | 600 | 86600 |



**Merits and Demerits of Arithmetic Mean.**

**Merits**

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. It is based upon all the observations.
4. It is amenable to algebraic treatment.

**Demerits**

1. It cannot be determined by inspection nor it can be located graphically.
2. Arithmetic mean cannot be used if we are dealing with qualitative characteristics which cannot be measured quantitatively, such as, intelligence, honesty, beauty, etc.
3. Arithmetic mean cannot be obtained if a single observation is missing or lost or is illegible unless we drop it out and compute the arithmetic mean of the remaining values.
4. Arithmetic mean is affected very much by extreme values. In case of extreme items, arithmetic mean gives a distorted picture of the distribution and no longer remains representative of the distribution.
5. Arithmetic mean may lead to wrong conclusions if the details of the data from which it is computed are not given.

**Weighted Mean**

Weighted mean enables us to calculate an average that takes into account the importance of each value to the overall total.

**Example**

|  |  |  |
| --- | --- | --- |
| **Grade of labor** | **Wage / hour (Rs)** | **No. of hours required** |
| **Unskilled** | 5 | 1 |
| **Semi-skilled** | 7 | 2 |
| **Skilled** | 9 | 5 |

Consider a company which uses three different grades of labor- unskilled, semiskilled and skilled as described in the above table to produce a product. If the company wants to know the average cost of labor per hour for the product a simple arithmetic average gives us

= = (5+7+9)/3=21/3=7

Using this average if we calculate the labor cost of 1 unit for the product to be 7(1+2+5)=56 but this answer is incorrect to be correct the answer must be calculated in the following way (5\*1)+(7\*2)+(9\*5)=64 and since there are eight hours of labor input the average labour cost per hour is 64/8=8.

Thus, we see that the weighted average give the correct values for the average hourly labour cost because they take into account that different amounts of each grade of labour are used in the product.

**Weighted Mean Formula**

Weighted mean is an average calculated to taking into account the importance of each value to the overall total, that is an average in which each observation value is weighted by some index of its importance?

Formula to calculate weighted mean =w=

x = each observation

w = weight assigned to each observation

∑w= sum of all weights.

**Harmonic Mean**

If x1, x2, x3,…xn  are n given observations then their H.M is given by

H=

Where N = Total frequency

X = value of variable or mid value of the class(in case of grouped or continuous distribution)

f is the corresponding frequency of X.

**Observations**: Since the reciprocals of the variables are involved, it gives greater weightage to smaller observations and as such is not very much affected by one or two big observations. It is not much affected by fluctuations of sampling. It is particularly useful in averaging special types of rates of ratios where time factor is variable, and the act being performed remains constant.

For example, a cyclist pedals from his house to his college at a speed of 10 kmph and back to home from the college at 15 kmph. Find the average speed.

Let the distance from the house to the college be X kms.

Going to the college X kms is covered in X/10 hrs.

while coming home, X kms is covered in X/15.

Therefore, the total distance of 2X kms is covered in (X/10 + X/15) hrs

Average speed = Total distance travelled



Total time taken



**Relationship between** A.M, G.M and H.M

A.M ≥ G.M ≥ H.M

**Problems**

1. A popular home furnishing shop ran 6 local TV channel advertisements during first week of May 2019. The following frequency distribution resulted.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No. of times Watched ad | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Frequency | 897 | 1082 | 1325 | 814 | 307 | 353 | 198 |

What is the average number of times the subscriber saw the advertisement during the period?

1. A professor has decided weighted average in figuring final grades for his seminar students. The homework average count for 20% of student’s grade the midterm 25% the final 35% the term paper 10% and quizzes 10%. From the following data compute the final average for the five students in the seminar.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Student | Homework | Quizzes | Paper | Mid Term | Final |
| 1 | 85 | 89 | 94 | 87 | 90 |
| 2 | 78 | 84 | 88 | 91 | 92 |
| 3 | 94 | 88 | 93 | 86 | 89 |
| 4 | 82 | 79 | 88 | 84 | 93 |
| 5 | 95 | 90 | 92 | 82 | 88 |



**Geometric Mean (G.M)**

Sometimes when we are dealing quantities that change over period of time we need to know an average rate of change such an average growth rate over period of several years in such cases simple arithmetic mean is inappropriate because it gives wrong answers in such cases we use geometric mean.



G.M is a measure of central trendy used to measure the average rate of change or growth of some quantity, computed taking nth root of the product of ‘n’ values representing the change.

If x1, x2, x3……. xn are n given observations then their G.M is given by

**Problems**

Suppose 1000 Dollars are deposited in a bank for 5 years with the rate of interest as mentioned.

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Rate of Interest | Growth factor | Savings at the end of the year |
| 1 | 7% | 1.07 | 1070 |
| 2 | 8% | 1.08 | 1155.6 |
| 3 | 10% | 1.1 | 1271.16 |
| 4 | 12% | 1.12 | 1423.67 |
| 5 | 15% | 1.15 | 1637.25 |

The average of growth factor is (1.07+1.08+1.1+1.12+1.15)/5 = 5.52/5 =1.104

1000\*1.104\*1.104\*1.104\*1.104\*1.104 = 1640

Which is not actual value, but if we consider



= 1.1037

1000\*1.1037\* 1.1037\*1.1037\*1.1037\*1.1037 = 1637.2 which is closer to the actual value.

Ex: A popular company shown the following percentage increase in net worth over last five years. Calculate the GM.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2014 | 2015 | 2016 | 2017 | 2018 |
| 5% | 10.5% | 9% | 6% | 7.5% |

**Merits and demerits of G.M**

**Merits**

1. It is rigidly defined.
2. Based on all observations.
3. Suitable for further mathematical treatments.

**Demerits**

1. Difficult to comprehend and calculate.
2. If any one of the observations is 0 G.M becomes 0.
3. If any one of the observations is negative G.M becomes imaginary.

G.M is used in calculating the compound interest and inflation over period time. Often A.M and G.M will not be too far apart, but a small difference can affect a proper decision. G.M is best fit when you are calculating the average percentage change in some variable over time

**Median**

The word median is synonymous with “middle,” and the sample median is indeed the middle value once the observations are ordered from smallest to largest.

The sample median is obtained by first ordering the n observations from smallest to largest with any repeated values included so that every sample observation appears in the ordered list. Then the median is given by

Ordered value --- if n is odd

Average of () and +1) ordered values. --- If n is even

For ex: Consider the 5 observations: 35 12 40 8 60. Arranging them in an ascending order: 8 12 35 40 60. Since the number of observations are odd, the median is 35. i.e. 3rd item in the list.

In case of even number of observations, median is obtained as the arithmetic mean of the two middle observations after they are arranged in ascending or descending order.

For ex: 8 12 35 40 50 60. Median = ½ (35+40) = 37.5.

For ex: **Calculating the median in a frequency distribution** where the variable takes the values x1, x2…. xn with respective frequencies f1, f2…….fn **with ∑ f = N.** The median is the size of (N+1)/2 th item or observation. In this case, the use of cumulative frequency distribution facilitates the calculations.

The steps are,

* Prepare the less than cumulative frequency distribution.
* Find N/2.
* See the CF, just greater than N/2.
* The corresponding value of the variable gives median.

Ex: 8 coins were tossed together and the number of heads(X) resulting was noted. The operation was repeated 256 times and the frequency distribution of the number of heads is given below.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **X** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| **f** | 1 | 9 | 26 | 59 | 72 | 52 | 29 | 7 | 1 |

Calculate the median.

N= ∑f =256

N/2= 128

The c.f just greater than 128 is 167. And the value of 167 corresponding is 4. Its median number of heads is 4.



|  |  |  |
| --- | --- | --- |
| **X** | **f** | **Less than c.f** |
| 0 | 1 | 1 |
| 1 | 9 | 1+9=10 |
| 2 | 26 | 10+26=36 |
| 3 | 59 | 36+59=95 |
| 4 | 72 | 95+72=167 |
| 5 | 52 | 167+52=219 |
| 6 | 29 | 219+29=248 |
| 7 | 7 | 248+7=255 |
| 8 | 1 | 255+1=256 |

For ex: A supermarket compares prices charged for identical items in all of its food stores. Here are the prices charged by each store for a Kg of green grams.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 92 | 89 | 89.5 | 91 | 91.5 | 93 | 92 | 91.5 | 90.5 | 90 |

Calculate the median price per kg, calculate mean price per Kg which value is better measure of central tendency of this data.

**Calculating the median for continuous frequency distribution**

* Prepare the less than CF distribution
* Find N/2
* Identify CF just greater than N/2
* The corresponding class contains the median value and is called median class.
* The value of median is obtained by using the formula.

**Median=l +h/f (N/2-C)**

l = lower limit of the median class.

f= frequency of the median class.

h= the magnitude of the median class.

N = the total frequency.

**Merits and demerits of median**

**Merits**

1. It is rigidly defined.
2. Easy to understand.
3. Since it is a positional average it is not affected by extreme values so very useful in skewed distribution such as wages, income etc. in case of extreme observations median is better average to use than arithmetic mean since A.M is affected by extreme values.
4. It can be computed with a distribution of open-end class
5. Median is only average that can be used while dealing with qualitative characteristics which can be measured quantitatively but still can be arranged in ascending or descending order like extremely poor, poor, middle income, above middle income, high income groups.

**Demerits**

1. The median not suitable for mathematical treatment.
2. It is just positional average; the magnitude of the observations cannot be understood.

Further understanding of median

The median value of the variable which divides the group into two equal parts one part comprising all the values greater and the other, all the values less than median. Thus, the median of a distribution may be defined as that value of the variable which exceeds and is exceeded by same number of observations. It is the value such that the number of observations above it is equal to the number of observations below it. That is median is the middle point of data set.

A measure of location that divides data into two halves.

**Mode**

The mode is the value that repeated most often in the data set.

Mode is the value which occurs most frequently in set of observations and around which the other items of the set cluster densely.

According A.M Tuttle mode is the value which has the greatest frequency density in its immediate neighbourhood.

In the following statements

1. The average size of shoe sold is 7.
2. The average height of an Indians is 1.6m.
3. The average size of shirts sold in India is 38.
4. In an average in Hyderabad household spends Rs 5000 on grocery.
5. The average that is referred is mode, neither mean nor median.

In all the above cases, the average referred to is mode. Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely. In other words, mode is the value of the variable which is predominant in the series.

In the case of discrete frequency distribution mode is the value of corresponding to maximum frequency.

For example, in the following frequency distribution:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| F | 5 | 8 | 16 | 26 | 23 | 14 | 6 | 3 |

The value of x corresponding to the maximum frequency 26 is 4. Hence mode is 4.

**Calculating mode from grouped data.**

When data are grouped in frequency distribution, we should consider the mode is in the class with most items. i.e. a class with highest frequency and this class is called modal class. The mode is calculated using the formula

M=Lm +

Where Lm = lower limit of the modal class

d1= frequency of the modal class – frequency of the class directly below it.

d2= frequency of the modal class – frequency of the class directly above it.

W=width of modal class interval.

**For example**

Find the mode of the following distribution:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| interval: | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
| Frequency | 5 | 8 | 7 | 12 | 28 | 20 | 10 | 10 |

**Solution.** Here maximum frequency is 28. Thus, the class 40-50 is the modal class. The value of mode is given.by

W=10 Lm=40, d1=28-12, d2=28-20

M=Lm +

40+10\*(28-12) / (28-12+28-20)

≈ 46.67

**Merits and Demerits of Mode**

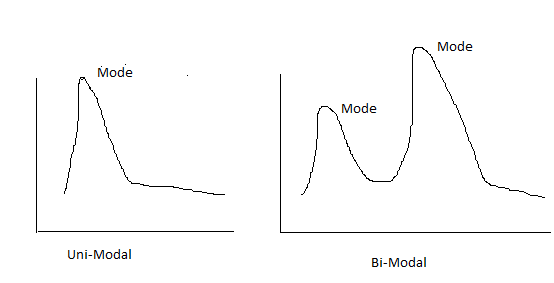
**Merits**

1. Mode is easy to understand. Sometimes it can be located just by inspection.
2. It is not affected by extreme values.
3. It can be used as a central location as qualitative and quantative data like median.
4. It can be used even when one or more classes are open ended.

**Demerits**

1. Mode is not well defined all the time. It is not always possible to find a clearly defined mode. In few cases, we may come across distributions with two modes. Such distributions are called bimodal. If a distribution has more than two modes, it is said to be multimodal.
2. Sometimes data set may not contain any repetitions in that case every value is a modal value which makes no sense.
3. It is not capable of further mathematical treatment.
4. As compared with mean, mode is affected to a greater extent by fluctuations of

sampling.



**Problems**

1. A librarian polled 20 different people as they left the library and asked them how many books they checked out the following are their responses

1 0 2 2 3 4 2 1 2 0 2 2 3 1 0 7 3 5 4 3

Compute the mode and mean, graph the data by plotting frequency versus number checked out. Is the mean or the mode a better measure of the central tendency of the data.

1. Marks of different students of sophomores in an engineering college.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | 47-51.9 | 52-56.9 | 57-67.9 | 62-62.9 | 67-71.9 | 72-76.9 | 77-81.9 |
| Frequency | 4 | 9 | 13 | 42 | 39 | 20 | 9 |

Calculate the modal value of the above distribution

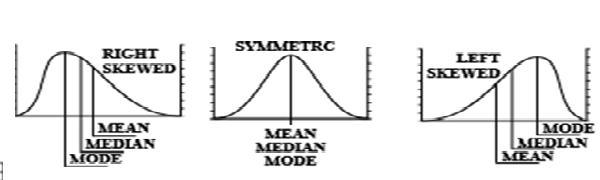
**Relationship between mean, median and mode**

If the distribution is symmetrical there is only one value for mean median and mode. In a positively skewed distribution (skewed to right) the mode is at the highest point of the distribution ,the median is to the right of that and the mean is to the right of both.

In negatively skewed distribution (skewed to the left) mode is the highest point of distribution the median is left of that and the mean is to the left of both.

When the population is skewed is negatively or positively the median is often the best measure of location because it is always between the mean and mode. But we observe that mode and mean are affected by the extreme values.

Graphical representation of mean median and mode.



Empirical relationship between mean median and mode

The following empirical relationship is given Prof. Karl Pearson

Mode=3 Median – 2Mean.

