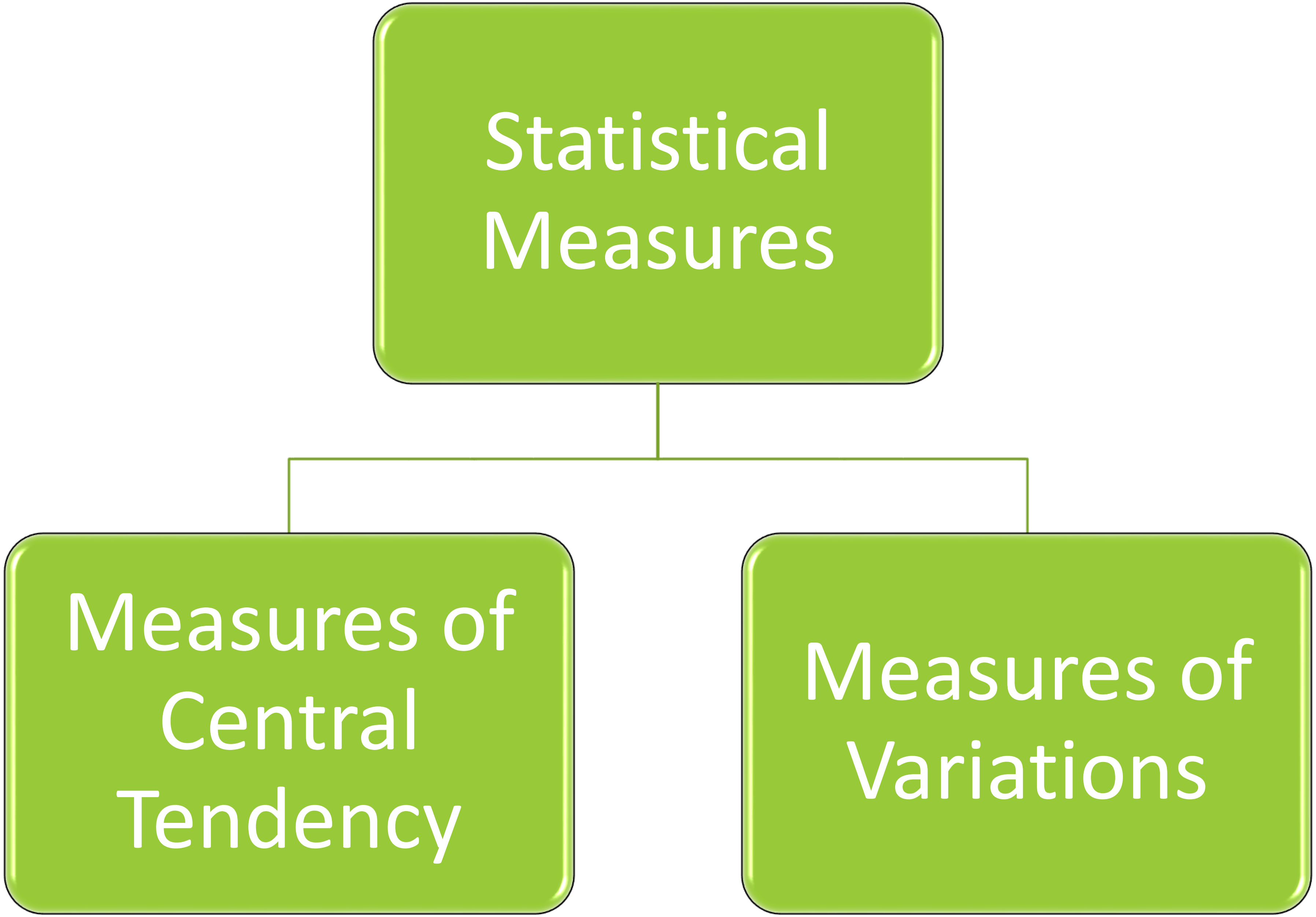
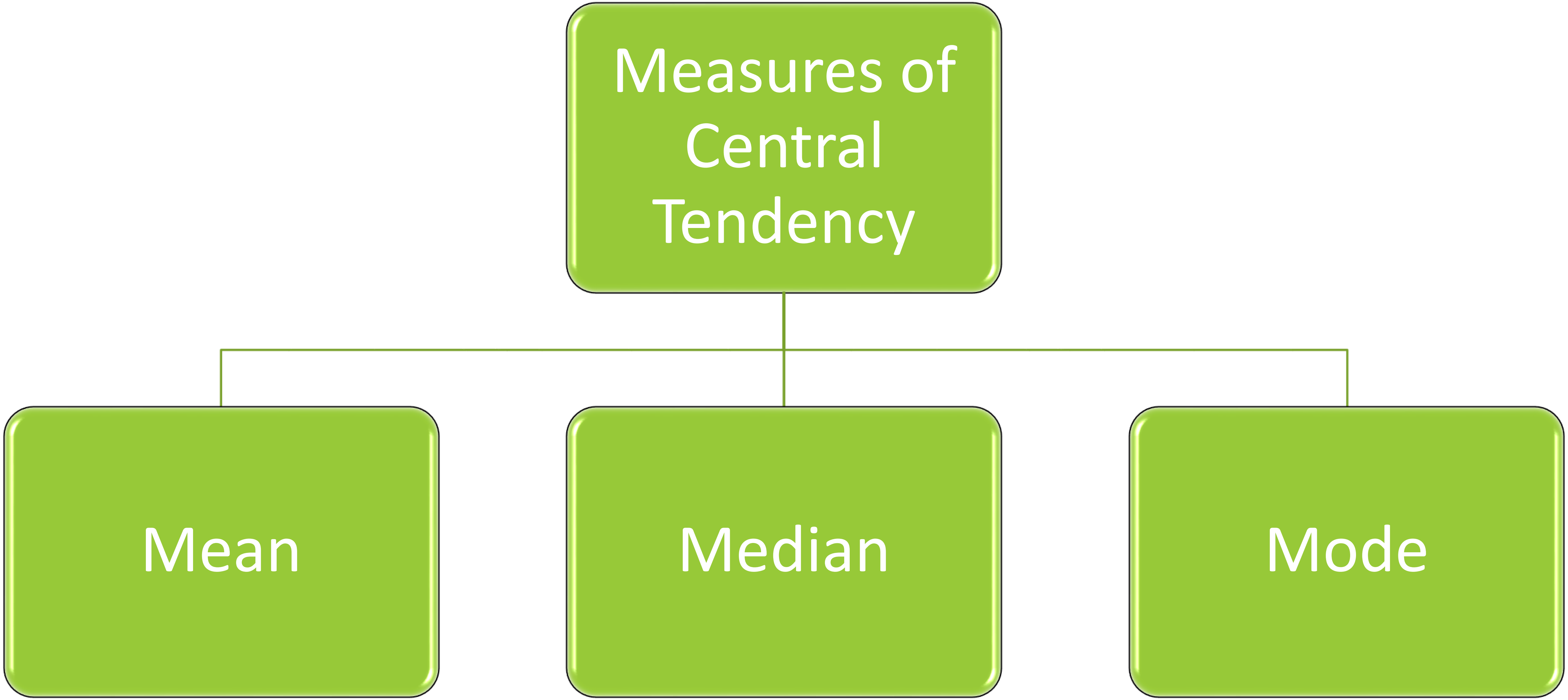
STATISTICAL MEASURES


Statistical Measures
Measures of Central Tendency
Measures of Variations

MEASURES OF CENTRAL TENDENCY

A measure of central tendency is a single value that attempts to describe a data by identifying the central position within that data. Measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics.


Measures of Central Tendency
Mean
Median
Mode

1. MEAN:

The mean (or average) is the most popular and well known measure of central tendency. It is used with both discrete and 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.

Mean for a sample is refer as *Sample mean* denoted by $\bar{x}$ (pronounced "x bar") and mean for a population is refer as *Population Mean* denoted by μ (pronounced “meo")

In general, we will discuss Sample mean through out our course.

We further classify it as for grouped and ungrouped data.

Tip: for small *n* we use simple formula and for large *n* use other formula

To proceed, we introduce the notation $x_1, x_2, ...x_i, ..., x_n$ for a general sample consisting of *n* measurements. Here x_1 represents the value of the first measurement, x_i is the i-th observation, and so on.

Now, Sample Mean for a small data or ungrouped data will be

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Sample Mean for a frequency distribution data or grouped data will be

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Example 1:

The final scores of 5 of my students for Calculus are 95, 83, 92, 81, 75. What is their mean score?

Solution:

As our sample size is small so, we can use the simple formula for finding the mean by sum up all the scores and divide by the total number of students.

Mean = $\bar{x}$ = (95+83+92+81+75)/5 = 85.2

Example 2: The following frequency distribution shows the heights (in inches) of 100 students in a class. Find the mean.

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{6558}{100} = 6558 \text{ inches}$$

Heights (inches)	f	x	fx
60 - 62	5	61	305
62 - 64	18	63	1134
64 - 66	42	65	2730
66 - 68	20	67	1340
68 - 70	8	69	552
70 - 72	7	71	497
Total	100		6558

2. MEDIAN:

As its name define that this is a synonym of middle therefore, it is the value that divides the distribution into two equal parts, so that half the cases are above it and half below it. We can define it as,

The median is the middle value, or average of middle values of a data. It is denoted by $\tilde{x}$.

We will discuss it first for ungrouped data and then for grouped data.

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,

$$\tilde{x} = \begin{cases} \text{The single middle value if } n \text{ is odd} & = \left(\dfrac{n+1}{2}\right)^{\text{th}} \text{ ordered value} \\ \text{The average of the two middle values if } n \text{ is even} & = \text{average of } \left(\dfrac{n}{2}\right)^{\text{th}} \text{ and } \left(\dfrac{n}{2}+1\right)^{\text{th}} \text{ ordered values} \end{cases}$$

EXAMPLE Calculation of the sample mean and median

In order to control costs, a company collects data on the weekly number of meals claimed on expense accounts. The numbers for five weeks are

15 14 2 7 and 13.

Find the mean and the median.

Solution The mean is

$$\overline{x} = \frac{15 + 14 + 2 + 27 + 13}{5} = 14.2 \quad \text{meals}$$

and, ordering the data from smallest to largest

$$2 \quad 13 \quad \underbrace{14} \quad 15 \quad 27$$

the median is the third largest value, namely, 14 meals.

Both the mean and median give essentially the same central value.

EXAMPLE Calculation of the sample median with even sample size

An engineering group receives email requests for technical information from sales and service persons. The daily numbers for six days are

11 9 17 19 4 and 15.

Find the mean and the median.

Solution The mean is

$$\overline{x} = \frac{11 + 9 + 17 + 19 + 4 + 15}{6} = 12.5 \text{ requests}$$

and, ordering the data from the smallest to largest

$$4 \quad 9 \quad \underbrace{11 \quad 15} \quad 17 \quad 19$$

the median, the mean of the third and fourth largest values, is 13 requests.

Median for frequency distribution (grouped data):

For frequency distribution, first we have to locate our median class. For that we find class boundaries to make our data continuous one and the cumulative frequencies to count our number of observations easily then we can determine our median class by finding where the $\frac{\sum f}{2}$th value lies in the data. The class interval is called median class.

Formula:

$$\tilde{x} = l + \frac{h}{f}\left(\frac{\sum f}{2} - C.f\right)$$

Where,

l = lower class boundary of median class

h = width of median class

f = frequency of median class

$\sum f$ = total frequency (n)

$C.f$ = cumulative frequency of preceding the median class

Example: For the frequency distribution given below, find median.

Class intervals	**2 -- 4**	**5 -- 7**	**8 – 10**	**11 – 13**	**14 – 16**
f	2	4	6	3	1

For given frequency distribution,

$\frac{\sum f}{2} = \frac{16}{2} = 8$th value is the median. Therefore, our required median will lie in 3rd class interval and this will our median class. Remember our median class will be the class whose cumulative frequency is greater or may be equals to median value.

C.I.	*C.B.*	*f*	*C.f.*
2 - 4	1.5 - 4.5	2	2
5 - 7	4.5 - 7.5	4	6
8 - 10	7.5 - 10.5	6	12
11 - 13	10.5 - 13.5	3	15
14 - 16	13.5 - 16.5	1	16
Total		16 = N	

8th value

For given frequency distribution, our median will be

$$\tilde{x} = l + \frac{h}{f}\left(\frac{\sum f}{2} - C.f\right)$$

$$\tilde{x} = 7.5 + \frac{3}{6}\left(\frac{16}{2} - 6\right)$$

$$\tilde{x} = 8.5$$

3. MODE:

Mode is the most frequent value occur in the data. One may define as the category or score with the largest frequency in the distribution.

It is not affected by mean or median. A data may not have any mode or may be have more than one modes. There is no notation for it. We can find the mode of an ungrouped data just by observing the values of data. While for grouped data we have a particular formula for the calculation.

Mode for frequency distribution:

For frequency distribution, first we have to locate our modal class. Modal is the class who has maximum frequency.

$$Mode = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

l = lower class boundary of modal class

h = width of modal class

f_m = maximum frequency

f_1 = frequency preceding to modal class

f_2 = frequency following to modal class

Example: For the same frequency distribution given before , find mode.

Class intervals	2 – 4	5 – 7	8 – 10	11 – 13	14 – 16
f	2	4	6	3	1
Class boundaries	1.5 – 4.5	4.5 – 7.5	7.5 – 10.5	10.5 – 13.5	13.5 – 16.5

$$Mode = 7.5 + 3\left(\frac{6-4}{2(6)-4-3}\right)$$

$$Mode = 8.7$$

Shape of the Distribution:

- ➢ Symmetric: (mean = median = mode)

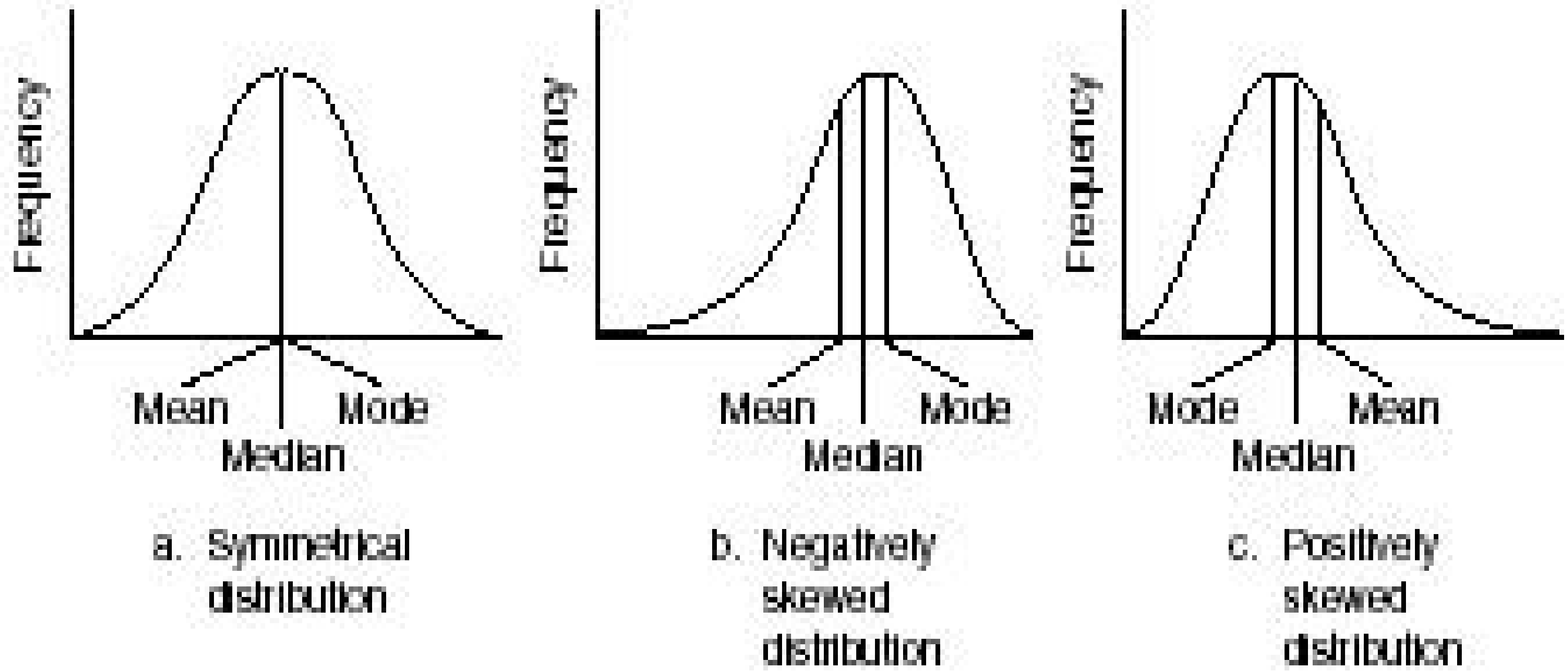
- ➢ Skew Symmetric:
  - Negatively Skewed (example: years of education)

    mean < median
  - Positively Skewed (example: income)

    mean > median

- ➢ Bimodal (two distinct modes)
- ➢ Multi-modal (more than 2 distinct modes)

Figure 4.6 Types of Frequency Distributions



Empirical relation between mean, median and mode for moderately skewed distribution is,

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

TASK:

Pick any data of your choice, may be the same data which you choose for last task and find its mean, median and mode. Also find that what type of relation they have.