Unit II

Central Tendency

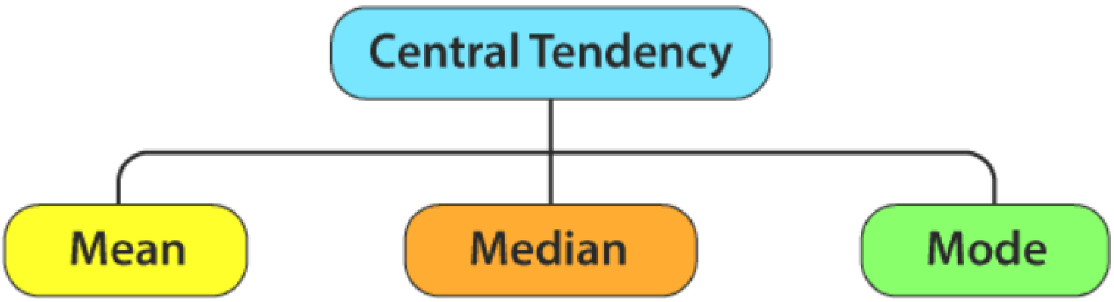
In statistics, the central tendency is the descriptive summary of a data set. Through the single value from the dataset, it reflects the centre of the data distribution. Moreover, it does not provide information regarding individual data from the dataset, where it gives a summary of the dataset. Generally, the central tendency of a dataset can be defined using some of the measures in statistics.

Definition

The central tendency is stated as the statistical measure that represents the single value of the entire distribution or a dataset. It aims to provide an accurate description of the entire data in the distribution.

Measures of Central Tendency

The central tendency of the dataset can be found out using the three important measures namely mean, median and mode.



Mean

The mean represents the average value of the dataset. It can be calculated as the sum of all the values in the dataset divided by the number of values. In general, it is considered as the arithmetic mean. Some other measures of mean used to find the central tendency are as follows:

- Geometric Mean
- Harmonic Mean
- Weighted Mean

It is observed that if all the values in the dataset are the same, then all geometric, arithmetic and harmonic mean values are the same. If there is variability in the data, then the mean value differs. Calculating the mean value is completely easy. The formula to calculate the mean value is given by:

$$Mean = \frac{x_1+x_2+..+x_n}{n}$$

The histogram given below shows that the mean value of symmetric continuous data and the skewed continuous data.

Median

Median is the middle value of the dataset in which the dataset is arranged in the ascending order or in descending order. When the dataset contains an even number of values, then the median value of the dataset can be found by taking the mean of the middle two values.

Consider the given dataset with the odd number of observations arranged in descending order – 23, 21, 18, 16, 15, 13, 12, 10, 9, 7, 6, 5, and 2

Median odd
23
21
18
16
15
13
12
10
9
7
6
5
2

Here 12 is the middle or median number that has 6 values above it and 6 values below it.

Now, consider another example with an even number of observations that are arranged in descending order – 40, 38, 35, 33, 32, 30, 29, 27, 26, 24, 23, 22, 19, and 17

	Median even
	40
	38
	35
	33
	32
	30
28	29
	27
	26
	24
	23
	22
	19
	17

When you look at the given dataset, the two middle values obtained are 27 and 29.

Now, find out the mean value for these two numbers.

i.e.,(27+29)/2 =28

Therefore, the median for the given data distribution is 28.

Mode

The mode represents the frequently occurring value in the dataset. Sometimes the dataset may contain multiple modes and in some cases, it does not contain any mode at all.

Consider the given dataset 5, 4, 2, 3, 2, 1, 5, 4, 5

Mode
5
5
5
4
4
3
2
2
1

Since the mode represents the most common value. Hence, the most frequently repeated value in the given dataset is 5.

What is a frequency distribution?

The **frequency** of a value is the number of times it occurs in a dataset. A **frequency distribution** is the pattern of frequencies of a variable. It's the number of times each possible value of a variable occurs in a dataset.

Types of frequency distributions

There are four types of frequency distributions:

- **Ungrouped frequency distributions:** The number of observations of each **value** of a variable.
  - You can use this type of frequency distribution for categorical variables.
- **Grouped frequency distributions:** The number of observations of each **class interval** of a variable. Class intervals are ordered groupings of a variable's values.
  - You can use this type of frequency distribution for quantitative variables.
- **Relative frequency distributions:** The proportion of observations of each value or class interval of a variable.
  - You can use this type of frequency distribution for **any type of variable** when you're more interested in **comparing frequencies** than the actual number of observations.
- **Cumulative frequency distributions:** The sum of the frequencies less than or equal to each value or class interval of a variable.
  - You can use this type of frequency distribution for **ordinal or quantitative variables** when you want to understand **how often observations fall below certain values**.

How to make a frequency table

Frequency distributions are often displayed using **frequency tables**. A frequency table is an effective way to summarize or organize a dataset. It's usually composed of two columns:

- The values or class intervals
- Their frequencies

The method for making a frequency table differs between the four types of frequency distributions. You can follow the guides below or use software such as Excel, SPSS, or R to make a frequency table.

How to make an ungrouped frequency table

1. **Create a table** with two columns and as many rows as there are values of the variable. Label the first column using the variable name and label the second column "Frequency." Enter the values in the first column.
    - For **ordinal variables**, the values should be ordered from smallest to largest in the table rows.
    - For **nominal variables**, the values can be in any order in the table. You may wish to order them alphabetically or in some other logical order.
2. **Count the frequencies.** The frequencies are the number of times each value occurs. Enter the frequencies in the second column of the table beside their corresponding values.
    - Especially if your dataset is large, it may help to count the frequencies by **tallying**. Add a third column called "Tally." As you read the observations, make a tick mark in the appropriate row of the tally column for each observation. Count the tally marks to determine the frequency.

Bird species	Tally	Frequency
Chickadee	III	3
Dove	I	1
Finch	IIII	4
Grackle	II	2
Sparrow	IIII	4
Starling	II	2

Scribbr

How to make a grouped frequency table

1. **Divide the variable into class intervals**. Below is one method to divide a variable into class intervals. Different methods will give different answers, but there's no agreement on the best method to calculate class intervals.
    - **Calculate the range.** Subtract the lowest value in the dataset from the highest.
    - **Decide the class interval width.** There are no firm rules on how to choose the width, but the following formula is a rule of thumb:

    $$\text{width} = \frac{\text{range}}{\sqrt{\text{sample size}}}$$

    You can round this value to a whole number or a number that's convenient to add (such as a multiple of 10).

- **Calculate the class intervals.** Each interval is defined by a lower limit and upper limit. Observations in a class interval are greater than or equal to the lower limit and less than the upper limit:

  $$\text{lower limit} \leq x < \text{upper limit}$$

  The lower limit of the first interval is the lowest value in the dataset. Add the class interval width to find the upper limit of the first interval and the lower limit of the second variable. Keep adding the interval width to calculate more class intervals until you exceed the highest value.

2. **Create a table** with two columns and as many rows as there are class intervals. Label the first column using the variable name and label the second column "Frequency." Enter the class intervals in the first column.
3. **Count the frequencies.** The frequencies are the number of observations in each class interval. You can count by tallying if you find it helpful. Enter the frequencies in the second column of the table beside their corresponding class intervals.
4. Example: Grouped frequency distributionA sociologist conducted a survey of 20 adults. She wants to report the frequency distribution of the ages of the survey respondents. The respondents were the following ages in years:

52, 34, 32, 29, 63, 40, 46, 54, 36, 36, 24, 19, 45, 20, 28, 29, 38, 33, 49, 37

5. $\text{range} = \text{highest} - \text{lowest}$
6. $\text{range} = 63 - 19$
7. $\text{range} = 44$
8. $\text{width} = \dfrac{\text{range}}{\sqrt{\text{sample size}}}$
9. $\text{width} = \dfrac{44}{\sqrt{20}}$
10. $\text{width} = 9.84$
11. Round the class interval width to 10.
12. The class intervals are $19 \leq a < 29$, $29 \leq a < 39$, $39 \leq a < 49$, $49 \leq a < 59$, and $59 \leq a < 69$.

Grouped frequency table of the ages of survey participants

Age, *a* (years)	Frequency
$19 \leq a < 29$	4
$29 \leq a < 39$	9
$39 \leq a < 49$	3
$49 \leq a < 59$	3
$59 \leq a < 69$	1

How to make a relative frequency table

1. **Create an ungrouped or grouped frequency table**.
2. **Add a third column to the table for the relative frequencies.** To calculate the relative frequencies, divide each frequency by the sample size. The sample size is the sum of the frequencies.

Relative frequency table of the frequency of bird species at a bird feeder

Bird species	Frequency	Relative frequency
Chickadee	3	$= \frac{3}{(3+1+4+2+4+2)}$ $= \frac{3}{16}$ $= .19$
Dove	1	.06
Finch	4	.25
Grackle	2	.13
Sparrow	4	.25
Starling	2	.13

Scribbr

How to make a cumulative frequency table

1. **Create an ungrouped or grouped frequency table** for an ordinal or quantitative variable. Cumulative frequencies don't make sense for nominal variables because the values have no order—one value isn't more than or less than another value.
2. **Add a third column to the table for the cumulative frequencies.** The cumulative frequency is the number of observations less than or equal to a certain value or class interval. To calculate the relative frequencies, add each frequency to the frequencies in the previous rows.
3. Optional: If you want to calculate the **cumulative relative frequency**, add another column and divide each cumulative frequency by the sample size.

Cumulative frequency table of the ages of survey participants

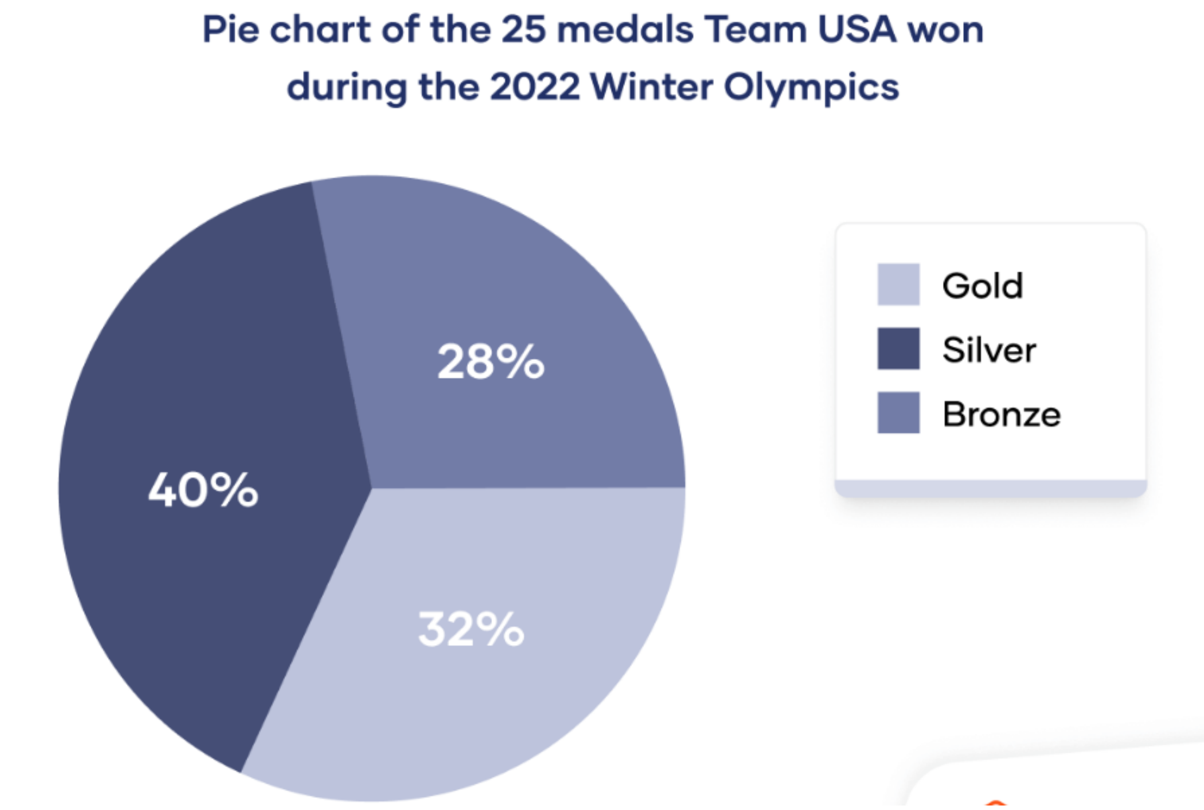
Age, *a* (years)	Frequency	Cumulative frequency	Cumulative relative frequency
$19 \le a < 29$	4	4	4 / 20 = .2
$29 \le a < 39$	9	9 + 4 = 13	.65
$39 \le a < 49$	3	9 + 4 + 3 = 16	.8
$49 \le a < 59$	3	19	.95
$59 \le a < 69$	1	20	1

How to graph a frequency distribution

Pie charts, bar charts, and histograms are all ways of graphing frequency distributions. The best choice depends on the type of variable and what you're trying to communicate.

Pie chart

A pie chart is a graph that shows the relative frequency distribution of a **nominal variable**.
A pie chart is a circle that's divided into one slice for each value. The size of the slices shows their relative frequency.
This type of graph can be a good choice when you want to emphasize that one variable is especially frequent or infrequent, or you want to present the overall composition of a variable.
A disadvantage of pie charts is that it's difficult to see small differences between frequencies. As a result, it's also not a good option if you want to compare the frequencies of different values.

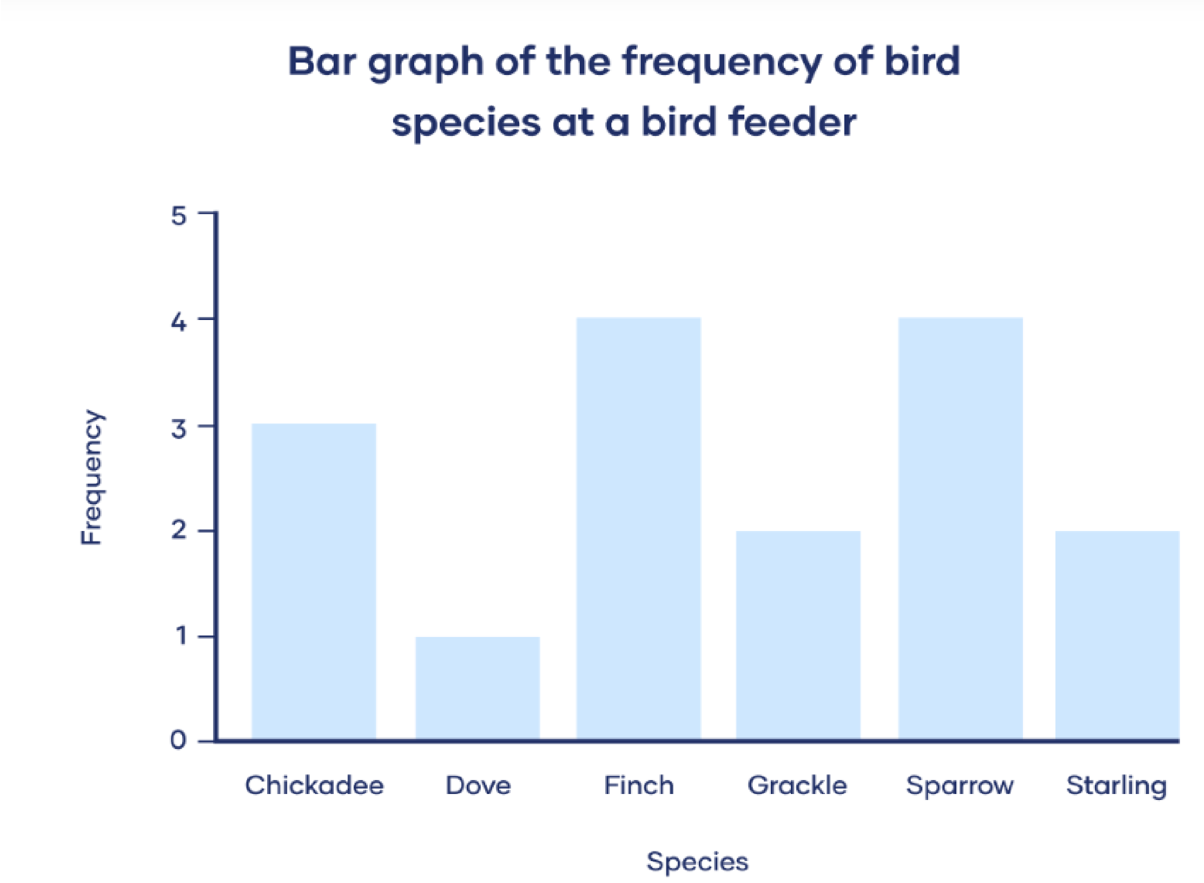


Bar chart

A bar chart is a graph that shows the frequency or relative frequency distribution of a **categorical variable** (nominal or ordinal).

The *y*-axis of the bars shows the frequencies or relative frequencies, and the *x*-axis shows the values. Each value is represented by a bar, and the length or height of the bar shows the frequency of the value.

A bar chart is a good choice when you want to compare the frequencies of different values. It's much easier to compare the heights of bars than the angles of pie chart slices.



Histogram

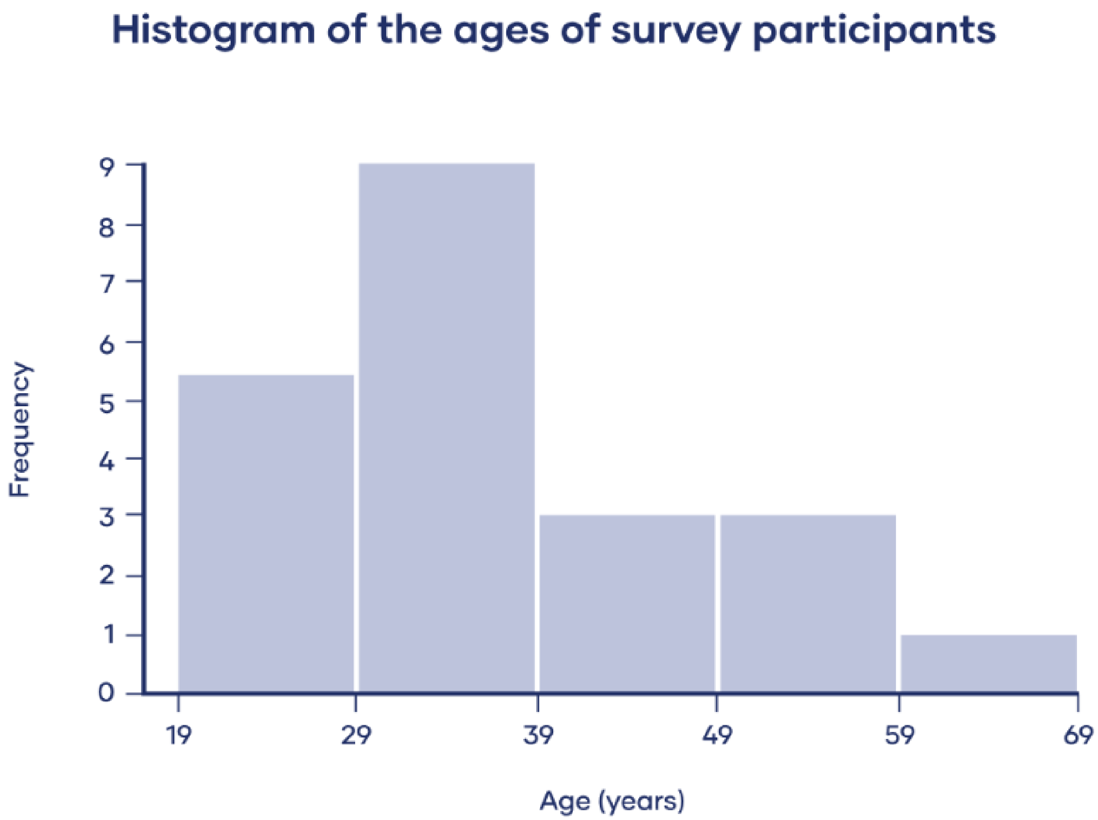
A histogram is a graph that shows the frequency or relative frequency distribution of a **quantitative variable**. It looks similar to a bar chart.

The continuous variable is grouped into **interval classes**, just like a grouped frequency table. The *y*-axis of the bars shows the frequencies or relative frequencies, and the *x*-axis shows the interval classes. Each interval class is represented by a bar, and the height of the bar shows the frequency or relative frequency of the interval class.

Although bar charts and histograms are similar, there are important differences:

	Bar chart	Histogram
Type of variable	Categorical	Quantitative
Value grouping	Ungrouped (values)	Grouped (interval classes)
Bar spacing	Can be a space between bars	Never a space between bars
Bar order	Can be in any order	Can only be ordered from lowest to highest

A histogram is an effective visual summary of several important characteristics of a variable. At a glance, you can see a variable's central tendency and variability, as well as what probability distribution it appears to follow, such as a normal, Poisson, or uniform distribution.



Continuous Frequency Distribution

A frequency distribution is a comprehensive way to organize raw data of a quantitative variable. It shows how different values of a variable are distributed and their corresponding frequencies. However, we can prepare two frequency distribution tables, namely discrete frequency distribution and continuous frequency distribution table. In this article, you will learn what is continuous frequency distribution, how to make a frequency table for a continuous variable with detailed steps and examples.

Continuous Frequency Distribution Definition

A continuous frequency distribution is a series in which the data are classified into different class intervals without gaps and their respective frequencies are assigned as per the class intervals and class width.

How to Calculate Frequency Distribution

Let's understand how to prepare a continuous frequency distribution here.

To prepare a frequency distribution, one should address the following five questions.

1. Should we have equal or unequal sized class intervals?

There are two situations where we can define unequal class interval sizes. They are:

(i) When we have data on income and other related variables where the range is very high.

(ii) If many values are concentrated in a small part of the range, using class intervals with equal sizes will lead to a loss of information on various values.

In all other cases, we can define class intervals of equivalent sizes in frequency distributions.

2. How many classes should we have?

Depending on the total number of observations, the number of classes could be between 6 and 15. Therefore, if we are using equal-sized class intervals, we can calculate the number of classes by dividing the range by the size of the class intervals.

3. What should be the size of each class?

We can determine the number of classes once we decide the class interval based on the range of the variable. Thus, we can notice that these two determinations are interlinked. Therefore, we cannot decide on one without deciding on the other.

4. How should we determine the class limits?

Class limits should be definite and explicitly stated. For example, we have two types of class intervals, such as:

(i) Exclusive class intervals: In this type of class interval, an observation equal to either the upper or the lower class limit is excluded from the frequency of the class.

(i) Inclusive class intervals: In this type of class interval, values equal to the lower and upper limits of a class are included in the frequency of the same class.

However, for discrete variables, we can use both exclusive and inclusive class intervals, whereas, for continuous variables, exclusive class intervals are used.

5. How should we get the frequency for each class?

We can find the frequency for each class by counting the number of values in a particular class.

The above points are necessarily to be followed in creating a continuous frequency distribution table.

A summary of the above defined process is given below:

Step 1: Determine the range of the data set.

Step 2: Divide the range by the number of the classes that we want our data in and then round up.

Step 3: Create class intervals using class width.

Step 4: Obtain the frequency for each class.

Let's apply the above process to get the frequency distribution table for the given data.

Continuous Frequency Distribution Table

Consider the following data.

17, 30, 37, 34, 39, 32, 30, 35, 12, 14, 12, 14, 14, 0, 25, 25, 25, 28, 47, 42, 49, 49, 45, 49, 46, 41, 60, 64, 62, 40, 43, 48, 48, 49, 49, 40, 41, 59, 51, 53, 82, 80, 85, 90, 98, 90,56, 55, 57, 55, 10, 14, 51, 50, 56, 70, 75, 64, 60, 66, 69, 62, 61, 70, 76, 70, 59, 56, 59, 57, 59, 55, 20, 22, 56, 51, 55, 56, 55, 50, 54, 66, 69, 64, 66, 60, 65, 62, 45, 47, 44, 40, 44, 65, 66, 65, 71, 82, 82, 90

The stepwise procedure to prepare a table of continuous frequency distribution is given below:

Step 1: Determine the range of the data set.

Maximum value = 98

Minimum value = 0

Range = Maximum value – Minimum value = 98 – 0 = 98

Step 2: Divide the range by the number of the classes that we want our data in and then round up.

Let the number of classes be 10.

Class width = 98/10 = 9.8

Thus, we can consider 10 as the class size.

Step 3: Create class intervals using class width.

For the above data, exclusive class intervals can be created and avoid taking the value which is equal to the upper limit of the class while writing the frequencies.

0 – 10, 10 – 20, 20 – 30, 30 – 40, 40 – 50, 50 – 60, 60 – 70, 70 – 80, 80 – 90, 90 – 100.

Step 4: Obtain the frequency for each class.

Class interval	Frequency	Corresponding observations
0 – 10	1	0
10 – 20	8	10, 12, 12, 14, 14, 14, 14, 17
20 – 30	6	20, 22, 25, 25, 25, 28
30 – 40	7	30, 30, 32, 34, 35, 37, 39
40 – 50	21	40, 40, 40, 41, 41, 42, 43, 44, 44, 45, 45, 46, 47, 47, 48, 48, 49, 49, 49, 49, 49
50 – 60	23	50, 50, 51, 51, 51, 53, 54, 55, 55, 55, 55, 55, 56, 56, 56, 56, 56, 57, 57, 59, 59, 59, 59
60 – 70	19	60, 60, 60, 61, 62, 62, 62, 64, 64, 64, 65, 65, 65, 66, 66, 66, 66, 69, 69
70 – 80	6	70, 70, 70, 71, 75, 76
80 – 90	5	80, 82, 82, 82, 85
90 – 100	4	90, 90, 90, 98
Total	100	

Continuous Frequency Distribution Examples

Example 1:

Prepare a frequency distribution by inclusive method taking a class interval of 7 from the following data.

28, 17, 15, 22, 29, 21, 23, 27, 18, 12, 7, 2, 9, 4, 1, 8, 3, 10, 5, 20, 16, 12, 8, 4, 33, 27, 21, 15, 3, 36, 27, 18, 9, 2, 4, 6, 32, 31, 29, 18, 14, 13, 15, 11, 9, 7, 1, 5, 37, 32, 28, 26, 24, 20, 19, 25, 19, 20, 6, 9

Solution:

For the given data:

Range = Maximum value – Minimum value = 37 – 1 = 36

Number of classes = 36/7 = 5.1 {since the class size is 7 as per the given}

Thus, we can define 5 classes.

The inclusive class intervals can be written as:

0 – 7, 8 – 15, 16 – 23, 24 – 31, 32 – 39

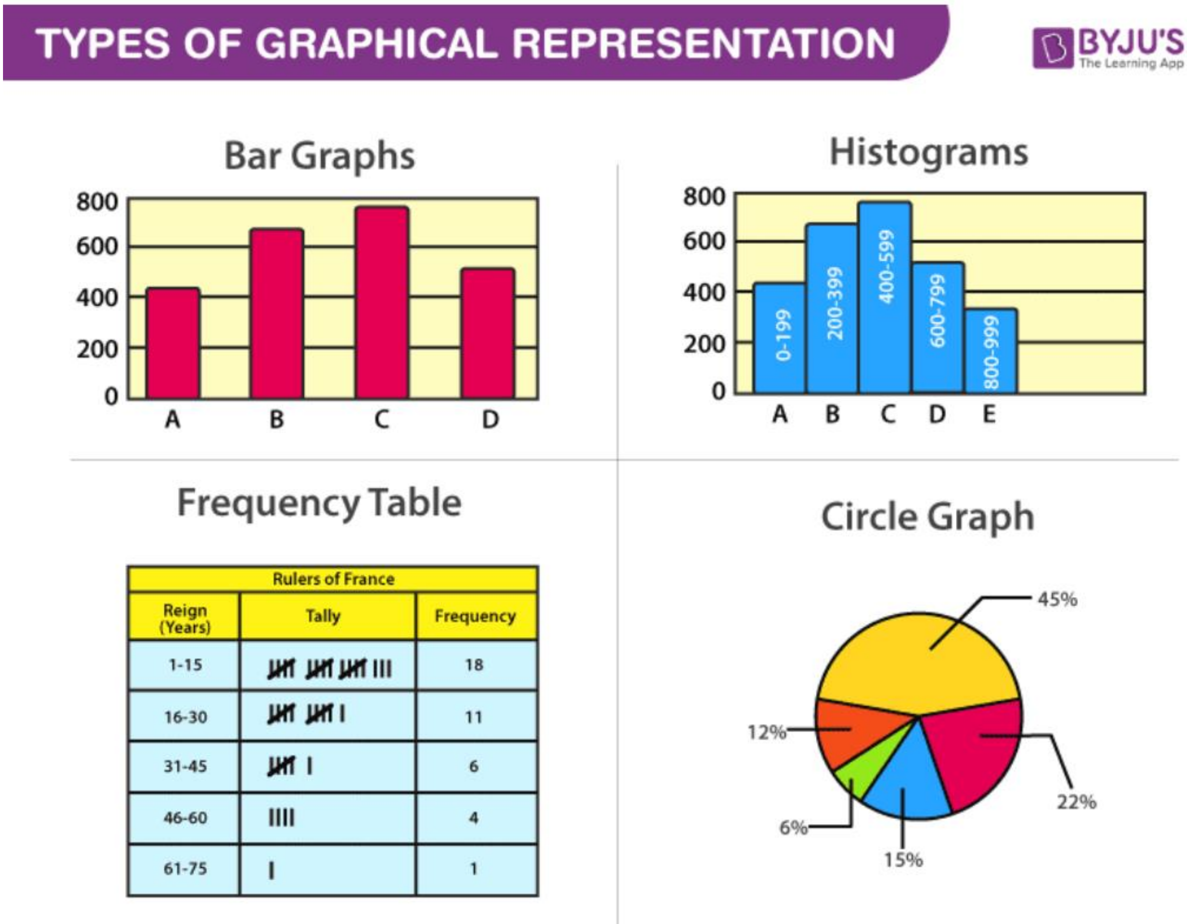
Let’s write the frequencies of these classes.

Class intervals	Frequencies	Corresponding data values
0 – 7	15	1, 1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 6, 6, 7, 7
8 – 15	15	8, 8, 9, 9, 9, 9, 10, 11, 12, 12, 13, 14, 15, 15, 15
16 – 23	14	16, 17, 18, 18, 18, 19, 19, 20, 20, 20, 21, 21, 22, 23
24 – 31	11	24, 25, 26, 27, 27, 27, 28, 28, 29, 29, 31
32 – 39	5	32, 32, 33, 36, 37
Total	60	

Graphical Representation

Graphical Representation is a way of analysing numerical data. It exhibits the relation between data, ideas, information and concepts in a diagram. It is easy to understand and it is one of the most important learning strategies. It always depends on the type of information in a particular domain. There are different types of graphical representation. Some of them are as follows:

- **Line Graphs** – Line graph or the linear graph is used to display the continuous data and it is useful for predicting future events over time.
- **Bar Graphs** – Bar Graph is used to display the category of data and it compares the data using solid bars to represent the quantities.
- **Histograms** – The graph that uses bars to represent the frequency of numerical data that are organised into intervals. Since all the intervals are equal and continuous, all the bars have the same width.
- **Line Plot –** It shows the frequency of data on a given number line. ‘ x ‘ is placed above a number line each time when that data occurs again.
- **Frequency Table** – The table shows the number of pieces of data that falls within the given interval.
- **Circle Graph** – Also known as the pie chart that shows the relationships of the parts of the whole. The circle is considered with 100% and the categories occupied is represented with that specific percentage like 15%, 56%, etc.
- **Stem and Leaf Plot –** In the stem and leaf plot, the data are organised from least value to the greatest value. The digits of the least place values from the leaves and the next place value digit forms the stems.
- **Box and Whisker Plot** – The plot diagram summarises the data by dividing into four parts. Box and whisker show the range (spread) and the middle (median) of the data.



Rulers of France		
Reign (Years)	Tally	Frequency
1-15	𝍸 𝍸 𝍸 III	18
16-30	𝍸 𝍸 I	11
31-45	𝍸 I	6
46-60	IIII	4
61-75	I	1

What are the Advantages of Graphical Method?

Some of the advantages of graphical representation are:

- It makes data more easily understandable.
- It saves time.
- It makes the comparison of data more efficient.

Mean Median Mode Formula

The Mean, Median and Mode are the three measures of central tendency. Mean is the arithmetic average of a data set. This is found by adding the numbers in a data set and dividing by the number of observations in the data set. The median is the middle number in a data set when the numbers are listed in either ascending or descending order. The mode is the value that occurs the most often in a data set and the range is the difference between the highest and lowest values in a data set.

The Mean

$$x = \frac{\sum x}{N}$$

Here,
∑ represents the summation

X represents observations
N represents the number of observations .

In the case where the data is presented in a tabular form, the following formula is used to compute the mean

Mean = ∑f x / ∑f

Where ∑f = N

The Median

If the total number of observations (n) is an odd number, then the formula is given below:

$$Median = \left(\frac{n+1}{2}\right)^{th} observation$$

If the total number of the observations (n) is an even number, then the formula is given below:

$$Median = \frac{\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n}{2}+1\right)^{th} observation}{2}$$

Consider the case where the data is continuous and presented in the form of a frequency distribution, the median formula is as follows.

Find the median class, the total count of observations ∑f.

The median class consists of the class in which (n / 2) is present.

$$\text{Median } = 1 + \left[\frac{\frac{\text{n}}{2} - \text{c}}{\text{f}}\right] \times \text{h}$$

Here

l = lesser limit belonging to the median class

c = cumulative frequency value of the class before the median class

f = frequency possessed by the median class

h = size of the class

The Mode

Consider the case where the data is continuous and the value of mode can be computed using the following steps.

a] Determine the modal class that is the class possessing the maximum frequency.

b] Calculate the mode using the below formula

$$\text{Mode } = 1 + \left[\frac{f_m - f_1}{2f_m - f_1 - f_2}\right] \times h$$

l = lesser limit of modal class

fm= frequency possessed by the modal class
f1= frequency possessed by the class before the modal class
f2= frequency possessed by the class after the modal class
h = width of the class

Geometric and Harmonic Mean

The geometric mean (G.M.) and the harmonic mean (H.M.) forms an important measure of the central tendency of data. They tell us about the central value of the data about which all the set of values of data lies. Suppose we have a huge data set and we want to know about the central tendency of this data set.

We have so many ways by which we can do so. But what if the data sets are fluctuating or we need to add or remove some of the data value? Calculating the average value or the central value will be a tiresome and troublesome task. So, we use geometric and harmonic means as our rescuer.

Geometric Mean

A geometric mean is a mean or average which shows the central tendency of a set of numbers by using the product of their values. For a set of n observations, a geometric mean is the nth root of their product. The geometric mean G.M., for a set of numbers $x_1, x_2, \dots, x_n$ is given as

$G.M. = (x_1 . x_2 \dots x_n)^{1/n}$

or, $G.M. = (\pi_{i=1}^{n} x_i)^{1/n} = {}^{n}\sqrt{(x_1, x_2, \dots, x_n)}$.

The geometric mean of two numbers, say x, and y is the square root of their product x×y. For three numbers, it will be the cube root of their products i.e., $(x\ y\ z)^{1/3}$.

Geometric Mean of Frequency Distribution

For a grouped frequency distribution, the geometric mean G.M. is

$G.M. = (x_1^{f1} . x_2^{f2} \dots x_n^{fn})^{1/N}$, where $N = \sum_{i=1}^{n} f_i$

Taking logarithms on both sides, we get

$\log G.M. = 1/N (f_1 \log x_1 + f_2 \log x_2 + \dots + f_n \log x_n) = 1/N [\sum_{i=1}^{n} f_i \log x_i]$.

Properties of Geometric Means

- The logarithm of geometric mean is the arithmetic mean of the logarithms of given values
- If all the observations assumed by a variable are constants, say K >0, then the G.M. of the observation is also K
- The geometric mean of the ratio of two variables is the ratio of the geometric means of the two variables
- The geometric mean of the product of two variables is the product of their geometric means

Geometric Mean of a Combined Group

Suppose G_1, and G_2 are the geometric means of two series of sizes n_1, and n_2 respectively. The geometric mean G, of the combined groups, is:

$\log G = (n_1 \log G_1 + n_2 \log G_2) / (n_1 + n_2)$

or, $G = \text{antilog}\ [(\log G_1 + n_2 \log G2) / (n_1 + n_2)]$

In general for n_i geometric means, $i = 1$ to k, we have

$G = \text{antilog}\ [(\log G_1 + n_2 \log G_2 + \ldots + n_k \log G_k) / (n_1 + n_2 + \ldots + n_k)]$

Advantages of Geometric Mean

- A geometric mean is based upon all the observations
- It is rigidly defined
- The fluctuations of the observations do not affect the geometric mean
- It gives more weight to small items

Disadvantages of Geometric Mean

- A geometric mean is not easily understandable by a non-mathematical person
- If any of the observations is zero, the geometric mean becomes zero
- If any of the observation is negative, the geometric mean becomes imaginary

Harmonic Mean

A simple way to define a harmonic mean is to call it the reciprocal of the arithmetic mean of the reciprocals of the observations. The most important criteria for it is that none of the observations should be zero.

A harmonic mean is used in averaging of ratios. The most common examples of ratios are that of speed and time, cost and unit of material, work and time etc. The harmonic mean (H.M.) of n observations is

$\text{H.M.} = 1 \div (1/n \sum_{i=1}^{n} (1/x_i))$

In the case of frequency distribution, a harmonic mean is given by

$\text{H.M.} = 1 \div [1/N (\sum_{i=1}^{n} (f_i / x_i)]$, where $N = \sum_{i=1}^{n} f_i$

Properties of Harmonic Mean

- If all the observation taken by a variable are constants, say k, then the harmonic mean of the observations is also k
- The harmonic mean has the least value when compared to the geometric mean and the arithmetic mean

$$\text{A.M.} \geq \text{G.M.} \geq \text{H.M.}$$

Advantages of Harmonic Mean

- A harmonic mean is rigidly defined
- It is based upon all the observations

- The fluctuations of the observations do not affect the harmonic mean
- More weight is given to smaller items

Disadvantages of Harmonic Mean

- Not easily understandable
- Difficult to compute

Weighted Average

Weighted average is considered the average where a weight is assigned to each of the quantities that are needed to be averaged. This weighting helps us in determining the respective importance of each quantity, on average. A weighted average can be considered to be more accurate than any simple average, as all the numbers in the set of data are assigned with identical weights. Let us explore the topic of weighted average, by understanding what is the meaning of weighted average, real-life examples, and solve a few examples using the formula.

What is the Meaning of Weighted Average?

Weight average also called weighted mean is helpful to make a decision when there are many factors to consider and evaluate. Each of the factors is assigned some weights based on their level of importance, and then the weighted average is calculated using a mathematical formula. The weighted average assigns certain weights to each of the individual quantities. The weights do not have any physical units and are only numbers expressed in percentages, decimals, or integers. The weighted average formula is the summation of the product of weights and quantities, divided by the summation of weights.

$$\text{Weighted Average} = \frac{\sum(\text{Weights} \times \text{Quantities})}{\sum \text{Weights}}$$

Definition of Weighted Average

When some quantities are more important than the others and do not contribute equally to the final result thus multiplying them to a coefficient is called weighted average. It is a simple process of deriving at an average value between two or quantities when weight is added to it. For example, a student realizes that the scores after an exam is two times more important than the scores acquired during the quiz. This is called the weighted average method.

Relationship amongst different Averages

In case of a moderately skewed distribution, the difference between mean and mode is almost equal to three times the difference between the mean and median. Thus, the empirical mean median mode relation is given as:

Mean – Mode = 3 (Mean – Median)

Or

Mode = 3 Median – 2 Mean
Either of these two ways of equations can be used as per the convenience since by expanding the first representation we get the second one as shown below:

Mean – Mode = 3 (Mean – Median)
Mean – Mode = 3 Mean – 3 Median
By rearranging the terms,
Mode = Mean – 3 Mean + 3 Median
Mode = 3 Median – 2 Mean

****