

Collection of Data and Formation of Frequency Distribution

Meaning of collection of Data

Collection of data is a statistical requirement. Statistics are a set or series of numerical data that acts as a facilitating factor of policy-making. In other words, numerical data establishes Statistics. Numerical data undergoes processing and manipulations before it aids the process of decision making. Hence, numerical data are the raw materials to statistics. These raw materials can originate from various sources. Statisticians and analysts collect these data in different methods.

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities, business, etc. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same.

Types of Data

Data may be qualitative or quantitative. Once you know the difference between them, you can know how to use them.

- 1. Qualitative Data:** They represent some characteristics or attributes. They depict descriptions that may be observed but cannot be computed or calculated. For example, data on attributes such as intelligence, honesty, wisdom, cleanliness, and creativity collected using the students of your class a sample would be classified as qualitative. They are more exploratory than conclusive in nature.
- 2. Quantitative Data:** These can be measured and not simply observed. They can be numerically represented and calculations can be performed on them. For example, data on the number of students playing different sports from your class gives an estimate of how many of the total students play which sport. This information is numerical and can be classified as quantitative.

Collection of Data

1. Primary Data:

These are the data which are collected from some primary sources i.e., a source of origin where the data generate.

These are collected for the first time by an investigator or an agency for any statistical analysis.

“Data which are gathered originally for a certain purpose are known as primary data.” — Horace Secrets

Merits:

1. It has high degree of accuracy.
2. For some enquiries, secondary data is not available.
3. These are more reliable.
4. It needs no extra precautions.

Demerits:

1. It requires lot of time.
2. It needs much money.
3. These data can be obtained through skilled persons only.
4. Sometimes, these data are not available altogether.

2. Secondary data:

These are the data which are collected from some secondary source i.e. the source of reservation storage where the data is collected by one person and used by other agency. These are collected as primary data and used by other as secondary data.

“The data which are used in an investigation, but which have been gathered originally by someone else for some other purpose are known as secondary data.” — Blair

Merits:

1. It is easy to collect.
2. Time and money is saved.
3. Sometimes primary data cannot be obtained.
4. Some data are more reliable than primary.

Demerits:

1. These are not reliable as primary data.

2. Extra caution is needed to use these data.
3. All types of data are not available.
4. Purpose of original collection may have been different.

Difference between Primary and Secondary Data

Data collection plays a very crucial role in the statistical analysis. In research, there are different methods used to gather information, all of which fall into two categories, i.e. primary data, and secondary data. As the name suggests, primary data is one which is collected for the first time by the researcher while secondary data is the data already collected or produced by others.

There are many differences between primary and secondary data, which are discussed in this article. But the most important difference is that primary data is factual and original whereas secondary data is just the analysis and interpretation of the primary data. While primary data is collected with an aim for getting solution to the problem at hand, secondary data is collected for other purposes.

BASIS FOR COMPARISON	PRIMARY DATA	SECONDARY DATA
Meaning	Primary data refers to the first hand data gathered by the researcher himself.	Secondary data means data collected by someone else earlier.
Data	Real time data	Past data
Process	Very involved	Quick and easy
Source	Surveys, observations, experiments, questionnaire, personal interview, etc.	Government publications, websites, books, journal articles, internal records etc.
Cost effectiveness	Expensive	Economical
Collection time	Long	Short
Specific	Always specific to the researcher's needs.	May or may not be specific to the researcher's need.
Available in	Crude form	Refined form

Accuracy and Reliability	More	Relatively less
---------------------------------	------	-----------------

Frequency Distribution

Frequency distribution is a representation, either in a graphical or tabular format that displays the number of observations within a given interval. The intervals must be mutually exclusive and exhaustive, and the interval size depends on the data being analyzed and the goals of the analyst. Frequency distributions are typically used within a statistical context.

As a statistical tool, a frequency distribution provides a visual representation for the distribution of a particular variable. Analysts often use it to show or illustrate the data collected in a sample. For example, the height of children can be split into several different categories or ranges. In measuring the height of 50 children, some are tall, and some are short, but there is a high probability of a higher frequency or concentration in the middle range. The most important factors are that the intervals used must be non-overlapping and must contain all of the possible observations.

Visual Representation

Frequency distributions can be presented as a frequency table, a histogram or a bar chart. Both histograms and bar charts provide a visual display using columns, with the y-axis representing the frequency count, and the x-axis representing the variable to be measured. In this example, the y-axis is the number of children, and the x-axis is the height. In general, the chart will show a normal distribution, which means that the majority of occurrences, or in this case children of a certain height, will fall in the middle column. In a histogram, the height of the column represents the range of values for that variable.

Frequency Distributions Used In Trading

Frequency distributions are not commonly used in the world of investments. However, traders who follow Richard D. Wyckoff, a pioneering trader in the early 20th century, use an approach to trading based on frequency distribution. Investment houses still use the approach, which

requires considerable practice, to teach traders. The frequency chart is referred to as a point-and-figure chart and was created out of a need for floor traders to take note of price action and to identify trends. The y-axis is the variable measured, and the x-axis is the frequency count. Each change in price action is denoted in X's and O's. Traders interpret it as an uptrend when three X's emerge; in this case, demand has overcome supply. In the reverse situation, when the chart shows three O's, it indicates that supply has overcome demand.

Example

Problem Statement:

Constructing a frequency distribution table of a survey was taken on Maple Avenue. In each of 20 homes, people were asked how many cars were registered to their households. The results were recorded as follows:

1 2 1 0 3 4 0 1 1 1 2 2 3 2 3 2 1 4 0 0

Solution:

Steps to be followed for present this data in a frequency distribution table.

1. Divide the results (x) into intervals, and then count the number of results in each interval. In this case, the intervals would be the number of households with no car (0), one car (1), two cars (2) and so forth.
2. Make a table with separate columns for the interval numbers (the number of cars per household), the tallied results, and the frequency of results in each interval. Label these columns Number of cars, Tally and Frequency.
3. Read the list of data from left to right and place a tally mark in the appropriate row. For example, the first result is a 1, so place a tally mark in the row beside where 1 appears in the interval column (Number of cars). The next result is a 2, so place a tally mark in the row beside the 2, and so on. When you reach your fifth tally mark, draw a tally line through the preceding four marks to make your final frequency calculations easier to read.

4. Add up the number of tally marks in each row and record them in the final column entitled Frequency.

Your frequency distribution table for this exercise should look like this:

Frequency table for the number of cars registered in each household		
Number of cars (x)	Tally	Frequency (f)
0		4
1		6
2		5
3		3
4		3

By looking at this frequency distribution table quickly, we can see that out of 20 households surveyed, 4 households had no cars, 6 households had 1 car.