

Classification and Tabulation

Classification is the process by which individuals and items are arranged into groups or classes according to their resemblance

Kinds of classification

1. Geographical Classification
2. Chronological Classification
3. Qualitative Classification
4. Quantitative Classification

Geographical Classification

Suppose the data are classified based on the Geographical location like country, state, district or town, the classification is called Geographical classification

Population (in millions) of different districts of Kerala

District	Population
Trivandrum	2938
Kollam	2398
Pathanamthita	1187
Allapuzha	1991
Kottayam	1819
Idikki	1077
Thrissur	2376

Chronological Classification or Classification based on Time

If the classification is done based on the period to which each value is associated (Days, months, years etc.). Such a classification is called Chronological Classification

Budgetary deficits of India from 1972-1981

Year	Deficit (in crores of Rs.)
1972	831
1973	650
1974	625
1975	490
1976	425
1977	975
1978	1590
1979	2700
1980	1975
1981	1539

Qualitative Classification

When the classification is done on the basis of qualitative characteristics like sex religion, colour etc. the classification is said to be qualitative

Type of classifications

- **Simple Classification-** Classification is based only on one qualitative Characteristic
- **Two way classification-** If the classification is done based on two characteristics the classification is called two way
- **Manifold classification-** Classification based on three or more qualitative characteristics

Quantitative Classification

Frequency distribution is used for the classification of numerical information collected about the quantitative data. The different type of frequency table are

- ❖ Discrete Frequency Distribution
- ❖ Continuous Frequency Distribution
- ❖ Cumulative Distributions
- ❖ Relative Frequency Distributions

Raw Data

Data collected in original form.

Frequency

The number of times a certain value or class of values occurs.

Frequency Distribution

The organization of raw data in table form with classes and frequencies.

Categorical Frequency Distribution

A frequency distribution in which the data is only nominal or ordinal.

Ungrouped Frequency Distribution

If the characteristic under consideration is discrete and the number of distinct values it takes is small, classification can be done by preparing a table giving the values taken by it and the number of units having those values for the characteristic. Such a table is called an ungrouped frequency table or an ungrouped frequency distribution.

The following are the number of flowers in 20 plants. Classify the data into an ungrouped frequency distribution.

2, 5, 3, 0, 3, 3, 5, 7, 2, 0, 2, 2, 3, 2, 5, 7, 5, 2, 3, 3

No of flowers	Frequency
0	2
2	6
3	6
5	4
7	2
Total	20

Grouped Frequency Distribution

A frequency distribution where several numbers are grouped into one class.

Exclusive and Inclusive classification

If the upper limit of one group is the lower limit of the next group the classification is called Exclusive classification. That is in this case the number of those who secured upper limit score is excluded from this class and is included into the next class.

Class	Frequency
0-10	2
10-20	6
20-30	6
30-40	4
40-50	2
Total	20

If the upper limit of one group is the number next to the lower limit of the preceding group the classification is called Inclusive classification. In this method the number of individuals securing lower limit and upper limit scores are included in the same table.

Class	Frequency
0-9	2
10-19	6
20-29	6
30-39	4
40-49	2
Total	20

Class Limits

Separate one class in a grouped frequency distribution from another. The limits could actually appear in the data and for each class we have one lower limit and upper limit. The class limits are those given in the grouped data.

Class Boundaries

Separate one class in a grouped frequency distribution from another. The boundaries have one more decimal place than the raw data and therefore do not appear in the data. There is no gap between the upper boundary of one class and the lower boundary of the next class. The lower class boundary is found by subtracting half the difference between the upper limit of a class and the lower limit of the succeeding class from the lower class limit and the upper class boundary is found by adding half the difference between the upper limit of a class and the lower limit of the succeeding class to the upper class limit.

For the exclusive classification both class limit and class boundaries are same. But for the inclusive classification class boundaries are to be found out. For the above examples class boundaries are -0.5-9.5, 9.5-

Class Width

The difference between the upper and lower boundaries of any class. The class width is also the difference between the lower limits of two consecutive classes or the upper limits of two consecutive classes. It is not the difference between the upper and lower limits of the same class.

Class Mark (Midpoint)

The number in the middle of the class. It is found by adding the upper and lower limits and dividing by two. It can also be found by adding the upper and lower boundaries and dividing by two.

Guidelines for classes

1. There should be between 5 and 20 classes.
2. The classes must be mutually exclusive. This means that no data value can fall into two different classes
3. The classes must be all inclusive or exhaustive. This means that all data values must be included.
4. The classes must be continuous. There are no gaps in a frequency distribution. Classes that have no values in them must be included (unless it's the first or last class which are dropped).
5. The classes must be equal in width. The exception here is the first or last class. It is possible to have a "below ..." or "... and above" class. This is often used with ages.

Creating a Grouped Frequency Distribution

1. Find the largest and smallest values
2. Compute the Range = Maximum - Minimum
3. Select the number of classes desired. This is usually between 5 and 20.
4. Find the class width by dividing the range by the number of classes and rounding up. There are two things to be careful of here. You must *round up*, not off. Normally 3.2 would round to be 3, but in rounding up, it becomes 4. If the range divided by the number of classes gives an integer value (no remainder), then you can either add one to the number of classes or add one to the class width.
5. Pick a suitable starting point less than or equal to the minimum value. You will be able to cover: "the class width times the number of classes" values. You need to cover one more value than the range. Follow this rule and you'll be okay: *The starting point plus the number of classes times the class width must be greater than the maximum value.* Your starting point is the lower limit of the first class. Continue to add the class width to this lower limit to get the rest of the lower limits.
6. To find the upper limit of the first class, subtract one from the lower limit of the second class. Then continue to add the class width to this upper limit to find the rest of the upper limits.
7. Find the boundaries by subtracting 0.5 units from the lower limits and adding 0.5 units from the upper limits. The boundaries are also half-way between the upper limit of one class and the lower limit of the next class. Depending on what you're trying to accomplish, it may not be necessary to find the boundaries.
8. Tally the data.
9. Find the frequencies.

Advantages and disadvantages of frequency table

Advantages

- ❖ It condenses the large mass of data into a comparatively small table.
- ❖ It attracts the attention of even a layman and gives him an insight into the nature of the observations.

- ❖ The general characteristics of the data like central tendency, scatter, symmetry etc. can be judged at least roughly from a frequency table.
- ❖ It helps further mathematical analysis of the data.
- ❖ It makes comparisons easier.

Disadvantages

- ❖ If the frequency distribution is grouped, the identity of the observations is lost.
- ❖ The selection of the class interval and lower bound of the first class are to a certain extent arbitrary. So different frequency tables into which the same data is classified may give contradictory impressions.

Relative Frequency

The frequency divided by the total frequency. This gives the percent of values falling in that class.

Class	Frequency	Relative frequency
0-10	2	0.1
10-20	6	0.3
20-30	6	0.3
30-40	4	0.2
40-50	2	0.1
Total	20	1.0

Cumulative Frequency

The cumulative frequency gives the cumulative frequency of each of the class. There are two types of cumulative frequency table namely

1. Less than cumulative frequency table
2. Greater than cumulative frequency table

To prepare the less than cumulative frequency table we write the upper bounds of the classes in the first column and the number of observations less than the each upper bound against it. These frequencies are called less than cumulative frequencies.

Similarly a more than cumulative frequency table gives the number of observations more than the lower limit of a class

Upper limit	Less than CF		Lower limit	Greater than CF
10	2		0	20
20	8		10	18
30	14		20	12
40	18		30	6
50	20		40	2

Tabulation

Tabulation is the presentation of classified data in a systematic and scientific way and in a form so that the special significance of data is clear. It is the process of presenting and arranging the data in an orderly manner in tables.

Advantages of presenting the data in a tabular form:

- ⇒ The data in a tabular form are presented economically.
- ⇒ The tabular method of expression is much clearer.
- ⇒ This form facilitates comparison, if any.
- ⇒ It summarizes the collected information in an attractive and impressive manner.

Rules for tabulation:

1. The table drawn should be attractive, impressive and clear. Thick and thin lines should be drawn to present the data with the true significance of its various sections.
2. Due prominence should be given to titles and sub-titles.
3. The titles should be simple and should give complete description in a clear and concise way.
4. The complicated tables should be avoided.
5. The rows and columns should be serialized.
6. The rows and columns should be have self explanatory title.
7. The units of measurement should be specified and defined.
8. The sources of material should be given clearly.
9. The table should be serially numbered.
10. A suitable title should be given to the table.
11. The table should be self explanatory.

Note:-Row headings are called stubs and column headings are called captions