

Classification

Meaning and Definition of Classification

Classification is the process of arranging the available facts into homogeneous groups or classes according to resemblances and similarities. Some important definitions are as follows:

“Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different but related parts” -- Secrist

“Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities, and giving expression to the unity of attributes that may subsist among a diversity of individuals”

-- R.L.Connor

Main features of classification

The main features of classification are as follows:

- The facts are classified into heterogeneous groups by the process of classification. All the units having similar characteristics are placed in one class or group.
- The basis of classification is unity in diversity.
- The classification may be either actual or notational
- It may be according to either attributes or

Objectives of Classification

- To condensed the mass data
- To facilitate comparison
- To clearly bring out the similarities and dissimilarities of data
- To prepare the basis for future analysis.
- To bring out relationship.

Main Rules of classification / elements of classification

1. Unambiguity or Exactness
2. Exhaustive and Mutually Exclusive
3. Stability
4. Flexibility
5. Homogeneity
6. Suitability
7. Arithmetical accuracy

Methods of Classification

Every item of the collected data has its own characteristics.

These characteristics can be of two types:

i) Descriptive: (eg., Honesty, beauty, sex etc.)

These characteristics are those which cannot be measured directly but they are counted on the basis of presence or absence. (Non-measurable characteristics or attributes)

(ii) Numerical: (eg., height, weight, profit etc.)

Conti...

Numerical facts are those which can be measured.

Statistical data can have two types of classification :

- 1) Classification according to Attributes or Qualitative classification.
- 2) Classification according to variables or Quantitative classification.

Classification according to Attributes or Qualitative Classification

- In this type of classification, data are classified on the basis of some attributes or quality like sex, literacy, religion, employment, etc. Such attributes can not be measured along with a scale. In this type we can only find out the presence or absence of the attributes, in the given units. Qualitative classification can be of two types:

Conti...

(i) Dichotomy or Two-fold Classification:- When on the basis of presence or absence of an attribute, the data are classified into two classes--- one possessing that attribute, and the other not possessing that attribute, it is called two-fold or dichotomous classification.

Conti...

For eg., if the students in a class are to be classified in respect of one attribute, say sex, then we can classify them into two classes namely that of males or females. Similarly they can also be classified into employed or unemployed on the basis of another attribute employment.

Conti...

(ii) Manifold Classification:- The classification, where two or more attributes are considered and several classes are formed, is called manifold classification.. First of all universe/population is divided onto two classes on the basis of one attribute, after that each class is further divided into two sub-classes on the basis of second attribute. If third attribute is also to be considered, each sub-class is further classified into two sub-classes.

Conti...

For eg., if we classify students simultaneously with respect to two attributes, i.e., sex and employment. Then students are first classified with respect to sex into males and females. Each of these classes may then be further classified into employed and unemployed on the basis of attribute employment and as such students are classified into four classes, namely, (i) male employed, (ii) female employed, (iii) male unemployed, (iv) female unemployed.

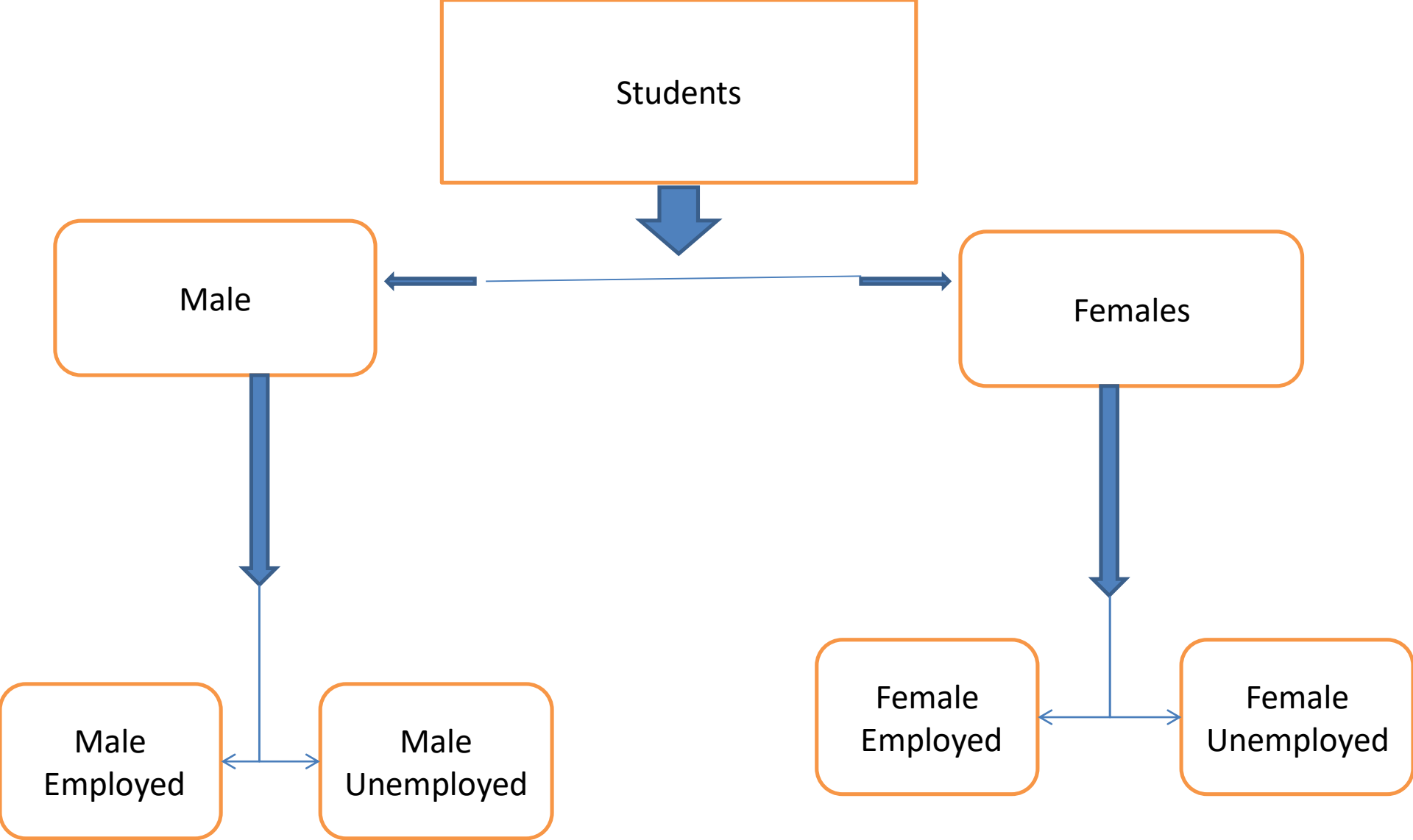
Conti...

For eg., if we classify students simultaneously with respect to two attributes, i.e., **sex and employment.**

Then students are first classified with respect to sex into males and females. Each of these classes may then be further classified into employed and unemployed on the basis of attribute employment and as such students are classified into four classes, namely, **(i) male employed, (ii) female employed, (iii) male unemployed, (iv) female unemployed.**

Conti...

Still, the classification may be further extended by considering other attributes like martial status etc. The above example of a manifold classification can also be explained by the following chart:



Quantitative Classification

If the data are classified on the basis of phenomena which is capable of quantitative measurement like age, height, weight, prices, production, income, expenditure, sales, profits, etc., it is termed as quantitative classification. The main methods of such classification are as follows:

Conti....

- (i) Geographical Classification
- (ii) Chronological Classification
- (iii) Variable Classification
 - (a) Continuous Variable (b) Discrete Variable

Conti...

(i) **Geographical Classification**: This type of classification is based on geographical or locational differences between various items in the data like states, cities, regions, zones etc. For eg. The yield of agricultural output per hectare for different countries in some given period may be presented as follows:

Agricultural Output of different countries (in Kg. per hectare)

Country	India	USA	Pakistan	Japan	china
Avg. Output	125	585	140	410	330

Conti...

(ii) **Chronological Classification**: When data are classified with respect to different periods of time (hour, day, week, month, year, etc.) it is known as chronological or temporal classification. For example, the population of India for different decades may be presented as follows:

Population of India (in Crores)

Year	1951	1961	1971	1981	1991	2000
Population	36.1	43.9	54.7	68.5	84.4	102.7

Conti...

i) **Variable Classification:** The quantitative phenomena under study is known as variable and classification on this basis is known as variable classification. Variables are of two kinds:

(a) Continuous variable (b) Discrete Variable

Continuous Variable: Those variables which can take all the possible values in a given specified range are termed as continuous variable.

Conti...

Discrete Variable: Those variables which cannot take all the possible values within a given specified range are termed as discrete (discontinuous) variables.

For example, the marks in a test (out of 100) of a group of students is a discrete variable since in this case marks can take only integral values from 0 to 100. It can not take all the values between 0 to 100.

Classification based on the basis of Discrete Values	
Height (cms.)	No. of Students
154	8
155	10
156	6
157	2
158	12
159	12
Total	50

Classification based on the basis of Continuous values	
Income (Rs.)	No. of Employees
1000-1500	15
1500-2000	33
2000-2500	22
2500-3000	18
3000-3500	12
Total	100

Technical terms in classification according to class-intervals

- Class Limits: The boundaries of the classes are called class limits.
- Class Frequency: The number of observations corresponding to a particular class is known as the frequency of the class.
- Magnitude of the class: The difference between the upper class limit and lower limit of a class is called the size or length of the class.

Conti..

- **Class Mid-Points**: The centre of the limits of a class is called the mid-point or mid-values.

Mid point of a class is defined as:

$$\frac{1}{2} \{ \text{Lower Class Limit} + \text{Upper Class limit} \}$$

Statistical Series

- If data are collected on the basis of characteristics which can be measured, such as age, income, height, etc. then the universe is classified on the basis of variables. This sort of classification is made in the form of statistical series.
- Statistical series are prepared to present the collected and classified data in a properly arranged way.

Kinds of Statistical Series

- Statistical series can be divided on the basis of their general character and construction. The following chart brings out clearly the various types of statistical series based on numerical or quantitative values.

Statistical Series

```
graph TD; A[Statistical Series] --> B[On the basis of general qualities or character]; A --> C[On the basis of construction]; B --> D["1. Time Series<br/>2. Spatial Series<br/>3. Condition series"]; C --> E[Univariate]; C --> F[Bivariate]; E --> G[Individual Series]; E --> H[Discrete Series]; E --> I[Continuous Series];
```

On the basis of general qualities or character

- 1. Time Series**
- 2. Spatial Series**
- 3. Condition series**

On the basis of construction

Univariate

Bivariate

Individual Series

Discrete Series

Continuous Series

Statistical Series based on General qualities

1. Time Series:- When data are represented with respect to different periods of time (hour, day, week, year, etc.) it is known as time-series.
2. Spatial Series:- When data are represented with respect to location or geographical area.
3. Condition Series:- If data are presented with reference to some condition, it is called condition series.

Statistical Series based on Construction **(structure)**

Basically, on the basis of construction statistical series
can be of two kinds:

- (i) Univariate Frequency distribution
- (ii) Bivariate Frequency Distribution

Univariate Frequency distribution

When frequencies are determined on the basis of one variable, it is known as univariate frequency distribution. It may be of the following three types:

(i) Series of Individual Observations

(ii) Discrete Series

(iii) Continuous series

(i) Series of Individual Observations

When the observations are expressed individually, the value of each and every item can be observed separately, it is known as individual series. Thus in this series values are equal to number of items.

- If data in an individual series are presented in an array, they will be called as organized data and if they are not in an array they will be known as raw or unorganized data.

(ii) Discrete Series

- Individual series, undoubtedly, make the data relatively easier.
- In a statistical data when the values are largely repeated then we can prepare a series showing the different values of the variables along with the number of repetitions of each value. Such a series is called a discrete series.

(iii) Continuous Series

- Continuous series is one where measurements are only approximations and are expressed in class intervals, i.e., within certain limits.
- In a continuous frequency distribution the class intervals theoretically continue from the beginning of the frequency distribution to the end without break.
- Some basic terms are:

Continuous Series...

- **Class:** Each stated interval, such as 10-20, is called a class.
- **Class Limit:** the boundaries of the class are called class-limit.
- **Class-Interval:** The width of the classes is called class interval.
- **Class frequencies:** The number of value of the series which falls in a class are known as class frequencies

Continuous Series...

Construction of Continuous series: Following steps are involved in the construction of continuous series:

1. Determining the Range
2. Determining the number of Classes
3. Determining the size of the class interval
4. Determination of type of class interval
5. Arrangement of Frequencies

Continuous Series...

1. Determining the Range: It is obtained by subtracting the size of the smallest item from the size of the largest item of the observed values. For example, in our illustration the smallest value is 432 and the highest value is 487. Thus the range will be $487 - 432 = 55$.
2. Determining the number of Classes: There are no hard and fast rules regarding the determination of

Continuous Series...

number of classes in any given observations.

In brief the following factors should be kept in consideration while determining the number of classes:

- i) The total frequency (i.e., total no. of observations)
- ii) The nature of the data, i.e., the size or magnitude of the values of the variable.
- iii) The accuracy aimed at.

Continuous Series...

3. Determination of the Size of the class intervals:

The difference between upper and lower limits of a class is called the magnitude of the class.

Generally the magnitude of the class interval depends on the range of the data and the number of classes.

Since the size of the class interval is inversely proportional to the number of classes in a given

Continuous Series...

4. **Determination of Types of Class Interval**: As we know that, each class is specified by two extreme values called the class limits, the smaller one being termed as lower limit and the larger one is the upper limit of the class. The classification of a frequency distribution into various is of following types:

(a) Exclusive Method (b) Inclusive Class Interval

Continuous Series...

a) Exclusive Method: Under this method lower limit is included and the upper limit is excluded while counting the observations. For instance, the class interval 10-20 includes all the values from 10 to 19 excluding the value 20 which is included in the next class interval 20-30.

In this method, the upper class limit of the preceding class will become the lower limit of the succeeding class.

Continuous Series...

The class intervals are formed continuously without any gap or break.

It is presumed that the values of the variable are both integers and fractional.

(b) Inclusive Class Interval: Under this method, both the limits are included while counting the observations. For instance, the class interval 10-19 includes all the values from 10 to 19, both inclusive

Continuous Series...

The next value 20 is included in the next class interval 20-29 and so on.

The fractional values between 19 and 20 cannot be accounted for in such a classification.

This method is used for those variables whose values are integers. Generally this method can not be adopted in the grouped series in which there are no fractional values of variable.

Continuous Series...

5. Arrangement of frequencies: Having defined the classes, the class intervals and the class-limits, we have to record the available data in the form of series for which we make use of tally-bars.

Each tally bar indicates the presence of one value in the class.

Tally bars are used in the form of 'Four and Cross Method'

Other Forms of Continuous series

a) Cumulative Frequency Series

(i) Less Than Cumulative Frequency Series

(ii) More Than Cumulative Series

(b) Open-End Classes Series

Other Forms of Continuous series...

a) Cumulative Frequency Series: Cumulative frequencies (c.f.) are derived by the cumulation of the frequencies of successive values.

Cumulative Frequency of a given variable or class represents the total frequency of a given variable or class represents the total frequency of all previous variables including the variable or the class. There are two types of cumulative frequencies series “less than” and “more than”

Other Forms of Continuous series...

(i) **Less than cumulative frequency series:** Less than c.f. for any value of the variable or class is obtained on adding successively the frequencies of all the previous values (or classes), including the frequency of variable (class) against which the totals are written , provided the values (classes) are arranged in ascending order of magnitude.

It is associated with the upper limits of the class.

C.I.	0-10	10-20	20-30	30-40	40-50
Frequency	4	5	12	9	15

Less Than Cumulative frequency

Upper Limits	c.f.
Less than 10	4
Less than 20	$(4+5)$ 9
Less than 30	$(9+12)$ 21
Less than 40	$(21+9)$ 30
Less than 50	$(30+15)$ 45

Other Forms of Continuous series...

(ii) **More than cumulative frequency series:** It is obtained similarly by finding the cumulative totals of frequencies starting from the highest value of the variable (class) to the lowest value (class).

It means we should start with the total frequency and go on deducting the frequencies for each value or class written against the value or class as more than value or lower limit of the class.

C.I.	0-10	10-20	20-30	30-40	40-50
Frequency	4	5	12	9	15

More Than Cumulative frequency

Lower Limits	c.f.
More than 0	45
More than 10	41
More than 20	36
More than 30	24
More than 40	15

?

Find Less than and More than c.f. ?

C.I.	0-10	10-20	20-30	30-40	40-50
Frequency	5	15	60	43	10

Other Forms of Continuous series...

(b) **Open-End Classes Series:** Under this series, the lower limit of the first class interval and the upper limit of the last class interval are not specified.

C.I.	Less than 10	10-20	20-30	30-40	More than 40
Frequency	5	15	60	43	10

Bivariate Frequency Distribution

- when data is classified with reference to two basis or criteria's simultaneously we can call such distributions as Bivariate distribution.
- **For example,** weights and heights of a group a group, ages of husbands and wives for a group of couples, marks in accountancy and statistics etc.
- Such type of distribution can be summarized in the form of a “two-way frequency table”. It is known as a “correlation table”

- In a Bivariate frequency distribution, the values of each variable are grouped into various classes keeping in view the same considerations as applied to a Univariate distribution.
- There will be class intervals for one set of values of variables as columns and the other set of values of variable as rows.

Case/Example

- Following figures give the ages of newly married husbands and their wives in years. Represent the following data by a frequency distribution.

Age of Husbands: 24,26,27,25,28,24,27,28,25,26,25,
26,27,25,27,26,25,26,26,26

Age of Wives: 17,18,19,17,20,18,18,19,18,19,17,
18,19,19,20,19,17,20,17,18

We have two variables – (i) age of husbands (ii) age of wives. Further we observe that age of husband's takes the values from 24 to 28 and age of wives takes the values from 17 to 20.

Age of Husband	Age of Wives				Total
	17	18	19	20	
24	I (1)	I (1)	---	---	2
25	III (3)	I (1)	I (1)	---	5
26	I (1)	III (3)	II (2)	I (1)	7
27	--	I (1)	II (2)	I (1)	4
28	--	--	I (1)	I (1)	2
Total	5	6	6	3	N = 20

TABLES

Table and Tabulation

- A table or statistical table is a systematic organization of data in columns and rows.
- “Tabulation” is an orderly arrangement of data in columns and rows systematically in a tabular form. It is the logical listing of related quantitative data in vertical columns and horizontal rows with explanatory and qualifying words, phrases and statements along with titles, heading and notes.

Table and Tabulation...

- It attempts to furnish the maximum information in the minimum possible space, without sacrificing the quality and usefulness of the data.
- It is an intermediate process between the collection of the data and the statistical analysis.
- “Tabulation involves the orderly and systematic presentation of numerical data in a form designed to illustrate the problem under consideration”-- Conner

Table and Tabulation...

- “A statistical table is a systematic organization of data in columns and rows” --- Prof. Neiswanger

Difference between Classification and Tabulation

- The data are classified before they are put in tables.
- The process of classification begins when the process of collection ends whereas the process of tabulation begins when the process of classification ends.
- In the process of classification data are arranged according to attributes and variables, whereas in the process of tabulation data are arranged in columns and rows.

Difference between Classification and Tabulation...

- The classification process refers to the arrangement of data into classes or groups, whereas the tabulation process refers to the placement of data in a tabular form. One refers to the **identification of data** and the other **presentation of data**.

Objectives/ functions of Tabulation

- To simplify complex data
- To present the data in a minimum space
- To facilitate the data for a comparative study
- To clarify the characteristics of data.

Importance or Advantages of Tabulation

- It simplifies complex data and the data presented are easily understood.
- It facilitates comparison of related facts.
- It facilitates computation of various statistical measures like averages, dispersion, correlation etc.
- It presents facts in minimum possible space and unnecessary repetitions and explanations are avoided.

Importance or Advantages of Tabulation...

- Required information can be easily located.
- Tabulated data are good for references and they make it easier to present the information in the form of graphs and diagrams.

Limitations of Tabulation

- A table contains only figures and not their description.
- It requires a specialized knowledge to understand a table. A layman cannot derive any conclusion from a table.
- A table does not lay emphasis on any section of particular importance.

Essentials or Characteristics of a Good table

1. Manageable Size
2. According to size
3. Comparable
4. Clarity
5. Attractive
6. Scientifically Prepared

Parts of a Table

1. Table Number:- A table should be numbered for easy reference and identification. This number, if possible should be written in the centre at the top of the table. Sometimes it is also written just before the title of the table.

2. Title of the Table:- A table should have a suitable title which is placed centrally on the top of a table just below the table number or just after table

Parts of a Table...

number in the same line in bold letters. The title describes the contents of the table briefly. As far as possible, the title should be complete and unambiguous as regards the subject matter of the data. It should be clear, properly worded and self-explanatory.

3. **Captions or column headings:-** A word or phrase which explains the contents of a column of a table

Parts of a Table...

is called the caption. A caption should be placed at the middle of the column. Under a caption, there may be subheads. When the items in the different columns are expressed in different units of measurement, the corresponding unit should be mentioned with the captions.

4. **Stubs or row Designations:-** Stubs stands for brief and self explanatory headings of horizontal rows.

Parts of a Table...

Normally, a relatively more important classification is given in rows. Also a variable with a large number of classes is usually represented in rows.

5. Head Notes (Prefatory Notes):- It is a statement given below the title which clarifies the contents of the table. It provides an explanation concerning the entire table or main parts of it.

6. Body:- The body of the table is the most important

Parts of a Table...

part of the table. The data is arranged in this part according to the description given by the captions and stubs.

7. **Footnotes:-** Footnotes are given at the foot of the table for explanation of any fact or information included in the table which needs some explanation. Thus they are meant for explaining or providing further details about the data, that have not been

Parts of a Table...

covered in title, captions or stubs.

8. **Source Note:-** It appears below the footnotes if they are used. The source notes are indicated, in case secondary data are used. These state the sources from which the data has been obtained. The source note should give the name of the author, table number, volume number, page number and the place of the publisher of the source from which data have been collected.

Rules for construction of Table

1. The table should be precise and simple so that the data presented should be easily understood. It should be free from irrelevant information, overlapping and ambiguities.
2. The table should contain all the important materials of the data.
3. The table should suit the size of the paper used. The width of the columns should be reasonable.

Rules for construction of Table...

Different types of lettering may be used to distinguish points of special emphasis.

4. Suitable title, captions and stubs are given to make the data more descriptive and self-explanatory.
5. The average, percentages, totals and grand totals are given both horizontally and vertically making approximations with no digital values.
6. All the items can be arranged either in an

Rules for construction of Table...

alphabetical, chronological or geographical order.

Sometimes, we may arrange them according to size, importance, emphasis or relationship.

7. Explanatory notes can be given wherever possible in the table. All the entries should be checked after they are made in the table. Over-writing should be avoided. Abbreviations should not be used.

Model Structure of a Table

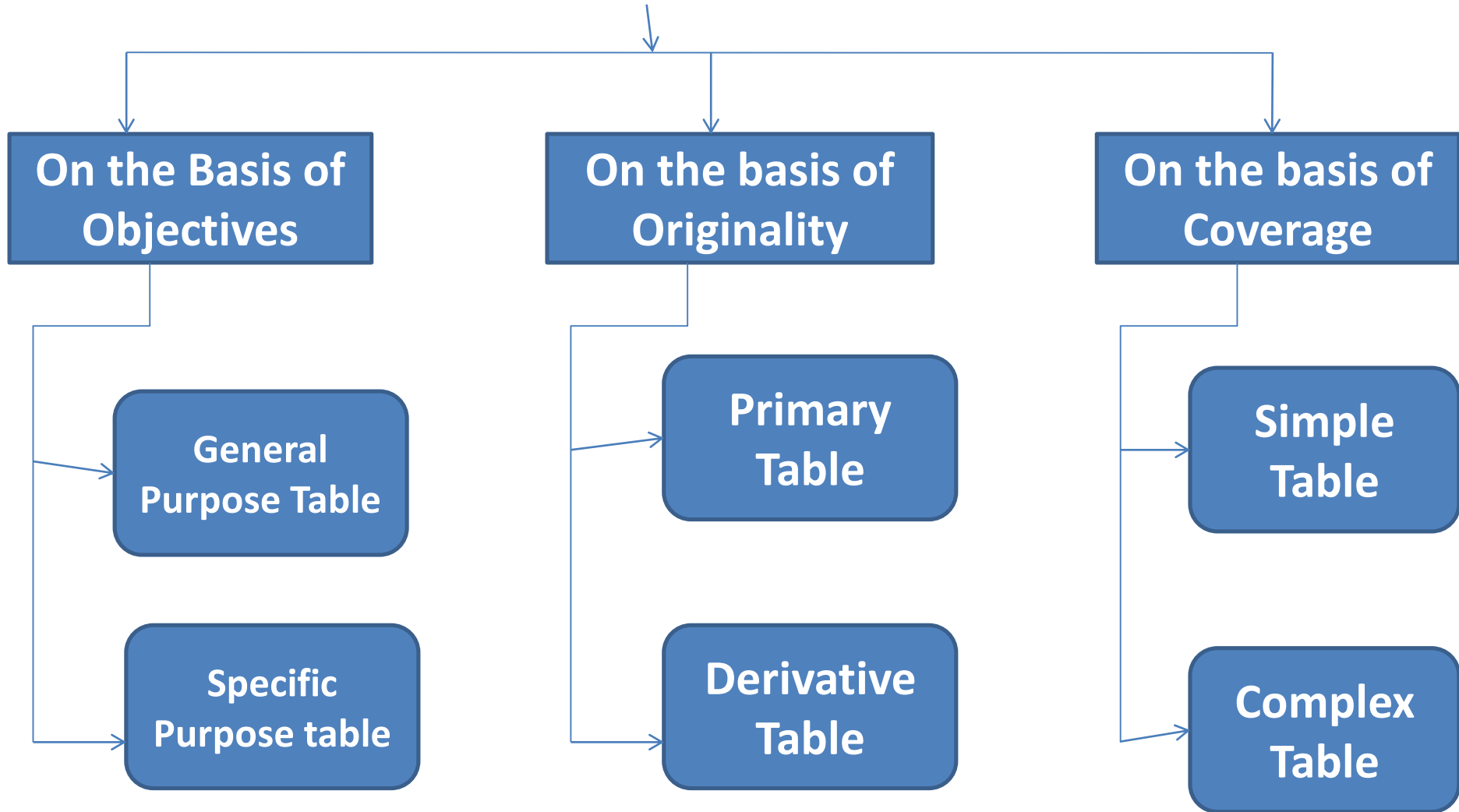
Table Number

Title of the Table

Sub Heading	Caption Headings	Total
	Caption Sub-Heading	
	Body	
Total		

Foot Note : 1-- 2—
Source Note : 1-- 2---

Kinds of Tables



On the Basis of Objective

General Purpose Tables: They are reference tables which are also called “information tables”. They are designed for use by research workers, statisticians and government agencies. They serve as a repository of information and are arranged for ease reference. They tell facts in general with no particular purpose.

Specific purpose tables: They are summary tables which are also called “interpretative tables”. They are analytical in nature and prepared with an idea of

On the Basis of Objective...

making comparative studies. They are derived from the general purpose tables to serve specific purpose. In such tables interpretative figures like ratios, percentages etc., are used in order to facilitate comparisons.

On the Basis of Originality

Primary Tables: The primary tables are those tables in which facts are expressed in the original form. They contain actual and absolute figures. They are also called “original tables”.

Derived or Derivative Tables: Derived table is one which contains figures and results derived from the original or primary data. It expresses the information in terms of ratios, percentages, aggregates, averages and other statistical measures.

On the Basis of Coverage...

Simple Tables: The simple tables are those tables which furnish information about only a single characteristic of the data. They are popularly known as “one-way tables”. The following is an example of simple table:

Table Showing the Number of Employees in Dena Bank according to Age Group	
Age in Years	No. of employees
Below 30	17
30-40	20
40-50	10
50 and above	3
Total	50

On the Basis of Coverage

Complex Tables: If the data are grouped on the basis of two or more characteristics or criterion simultaneously, then we get a complex table. The complex table may be of the following types

(a) Two-way Tables

(b) Three-way Tables

(c) Manifold Tables

On the Basis of Coverage...

(a) **Two-way Tables:-** They are called “doubled tables” which furnish information about two characteristics of the data. The two attributes of the data are interrelated for e.g., the number of students—male and female, male— married and unmarried, and female— married and unmarried.

(b) **Three-way Tables**

(c) **Manifold Tables**

Table Showing the Number of Employees in Dena Bank according to Age, Sex and Martial Status

<div> <div>Age in Years</div> <div>Sex</div> <div>→</div> </div>	No. of employees		Total
	Male	Female	
Below 30	10	7	17
30-40	11	9	20
40-50	3	7	10
50 and above	1	2	3
Total	25	25	50

On the Basis of Coverage...

(b) **Three-way Tables:-** The three way tables are those in which the data are classified with reference to three characteristics. They furnish information regarding three interrelated attributes of the same data. For e.g., the number of students—male and female, married and unmarried, and graduates and undergraduates. The following is an example of ‘Three Way table’:

Table Showing the Number of Employees in Dena Bank according to Age and Sex

<div>Sex</div> <div>Martial Status</div>	No. of employees						Total
	Male			Female			
Age Group	Married	Unmarr ied	Total	Married	Unmarr ied	Total	
Below 30	3	7	10	3	4	7	17
30-40	3	8	11	7	2	9	20
40-50	2	1	3	7	--	7	10
50 and above	1	--	1	2	--	2	3
Total	9	16	25	19	6	25	50

On the Basis of Coverage...

(c) **Manifold Tables:-** These are high order tables having more than three characteristics of the same data which are interrelated. Manifold or higher order tables are commonly used in presenting population census data.

Expected Questions

1. Define classification. Explain the purpose and methods of classification, giving suitable examples related to Business and Industry.
2. Distinguish between classification and tabulation.
3. Discuss in brief the modes of classification.
4. Discuss the purpose, methods and importance of classification.
5. What are the objectives of classification of data?

Expected Questions

6. Explain classification according to attributes and classification according to class-intervals with suitable examples.
7. Define tabulation. Describe different parts of a table .
What precautions would you observe in construction of a table.
8. What is a statistical table? Mention the rules of the construction of a table.