**INTRODUCTION TO STATISTICS**

**Meaning of Statistics**

The term statistics mean that the numerical statement as well as statistical methodology. When it is used in the sense of statistical data it refers to quantitative aspects of things and is a numerical description.

***Example***: Income of family, production of automobile industry, sales of cars etc. There quantities are numerical. But there are some quantities which are not in themselves numerical but can be made so by counting. The sex of a baby is not a number, but by counting the number of boys, we can associate a numerical description to sex of all new born babies, for an example, when saying that 60% of all live-born babies are boy. This information then comes within the realm of statistics.

**Definition**

The word statistics can be used is two senses, viz, singular and plural. In narrow sense and plural sense, statistics denotes some numerical data (statistical data). In a wide and singular sense statistics refers to the statistical methods. Therefore, these have been grouped under two heads – ‘Statistics as a data” and “Statistics as a methods”.

**Statistics as a Data**

Some definitions of statistics as a data are

1. Statistics are numerical statement of facts in any department of enquiring placed in relation to each other.

- Bowley

1. By statistics we mean quantities data affected to a marked extent by multiplasticity of course.

- Yule and Kendall

1. By statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for pre-determinated purpose and placed in relation to each other.

- H. Secrist

This definition is more comprehensive and exhaustive. It shows more light on characteristics of statistics and covers different aspects.

Some characteristics the statistics should possess by H. Secrist can be listed as follows.

* Statistics are aggregate of facts
* Statistics are affected to a marked extent by multiplicity of causes.
* Statistics are numerically expressed
* Statistics should be enumerated / estimated
* Statistics should be collected with reasonable standard of accuracy
* Statistics should be placed is relation to each other.

**Statistics as a methods**

***Definition***

1. “Statistics may be called to science of counting”

- A.L. Bowley

1. “Statistics is the science of estimates and probabilities”.

- Boddington

1. Dr. Croxton and Cowden have given a clear and concise definition.

“Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data”.

**According to Croxton and Cowden there are 5 stages.**

1. ***Collection of Data***

A structure of statistical investigation is based on a systematic collection of data. The data is classified into two groups

i) Internal data and

ii) External data

Internal data are obtained from internal records related to operations of business organization such as production, source of income and expenditure, inventory, purchases and accounts.

The external data are collected and purchased by external agencies. The external data could be either primary data or secondary data. The primary data are collected for first time and original, while secondary data are collected by published by some agencies.

1. ***Organisation of data***

The collected data is a large mass of figures that needs to be organised. The collected data must be edited to rectify for any omissions, irrelevant answers, and wrong computations. The edited data must be classified and tabulated to suit further analysis.

1. ***Presentation of data***

The large data that are collected cannot be understand and analysis easily and quickly. Therefore, collected data needs to be presented in tabular or graphic form. This systematic order and graphical presentation helps for further analysis.

1. ***Analysis of data***

The analysis requires establishing the relationship between one or more variables. Analysis of data includes condensation, abstracting, summarization, conclusion etc. With the help of statistical tools and techniques like measures of dispersion central tendency, correlation, variance analysis etc analysis can be done.

1. ***Interpretation of data***

The interpretation requires deep insight of the subject. Interpretation involves drawing the valid conclusions on the bases of the analysis of data. This work requires good experience and skill. This process is very important as conclusions of results is done based on interpretation.

We can define statistics as per Seligman as follows.

**“Statistics is a science which deals with the method and of collecting, classifying, presenting, comparing and interpretating the numerical data collected to throw light on enquiry”.**

**Importance of statistics**

In today’s context statistics is indispensable. As the use of statistics is extended to various field of experiments to draw valid conclusions, it is found increased importance and usage. The number of research investigations in the field of economics and commerce are largely statistical. Further, the importance and statistics in various fields are listed as below.

1. ***State Affairs*:** In state affairs, statistics is useful in following ways
2. To collect the information and study the economic condition of people in the states.
3. To assess the resources available in states.
4. To help state to take decision on accepting or rejecting its policy based on statistics.
5. To provide information and analysis on various factors of state like wealth, crimes, agriculture experts, education etc.
6. ***Economics*:** In economics, statistics is useful in following ways
7. Helps in formulation of economic laws and policies
8. Helps in studying economic problems
9. Helps in compiling the national income accounts.
10. Helps in economic planning.
11. ***Business***
12. Helps to take decisions on location and size
13. Helps to study demand and supply
14. Helps in forecasting and planning
15. Helps controlling the quality of the product or process
16. Helps in making marketing decisions
17. Helps for production, planning and inventory management.
18. Helps in business risk analysis
19. Helps in resource long term requirements, in estimating consumers preference and helps in business research.
20. ***Education*:** Statistics is necessary to formulate the polices regarding start of new courses, consideration of facilities available for proposed courses.
21. ***Accounts and Audits:***
22. Helps to study the correlation between profits and dividends enable to know trend of future profits.
23. In auditing sampling techniques are followed.

**Functions of statistics**

Some important functions of statistics are as follows

1. To collect and present facts in a systematic manner.
2. Helps in formulation and testing of hypothesis.
3. Helps in facilitating the comparison of data.
4. Helps in predicting future trends.
5. Helps to find the relationship between variable.
6. Simplifies the mass of complex data.
7. Help to formulate polices.
8. Helps Government to take decisions.

**Limitations of statistics**

1. Does not study qualitative phenomenon.
2. Does not deal with individual items.
3. Statistical results are true only on an average.
4. Statistical data should be uniform and homogeneous.
5. Statistical results depend on the accuracy of data.
6. Statistical conclusions are not universally true.
7. Statistical results can be interpreted only if person has sound knowledge of statistics.

**Distrust of Statistics**

Distrust of statistics are due to lack of knowledge and limitations of its uses, but not due to statistical sciences.

Distrust of statistics are due to following reasons.

* 1. Figures are manipulated or incompleted.
  2. Quoting figures without their context.
  3. Inconsistent definitions.
  4. Selection of non-representative statistical units.
  5. Inappropriate comparison
  6. Wrong inference drawn.
  7. Errors in data collection.

**Statistical Data**

Statistical investigation is a long and comprehensive process and requires systematic collection of data in large size. The validity and accuracy of the conclusion or results of the study depends upon how well the data were gathered. The quality of data will greatly influence the conclusions of the study and hence importance is to be given to the data collection process.

Statistical data may be classified as Primary Data and Secondary Data based on the sources of data collection.

* **Primary data**

Primary data are those which are collected for the first time by the investigator / researchers and are thus original in character. Thus, data collected by investigator may be for the specific purpose / study at hand. Primary data are usually in the shape of raw materials to which statistical methods are applied for the purpose of analysis and interpretation.

* **Secondary data**

Secondary have been already collected for the purpose other than the problem at hand. These data are those which have already been collected by some other persons and which have passed through the statistical analysis at least once. Secondary data are usually in the shape of finished products since they have been already treated statistically in one or the other form. After statistical treatment the primary data lose their original shape and becomes secondary data. Secondary data of one organisation become the primary data of other organization who first collect and publish them.

**Primary Vs Secondary Data**

* Primary data are originated by researcher for specific purpose / study at hand while secondary data have already been collected for purpose other than research work at hand.
* Primary data collection requires considerably more time, relatively expensive. While the secondary data are easily accessible, inexpensive and quickly obtained.

**Table – A compression of Primary and Secondary Data**

|  |  |  |
| --- | --- | --- |
|  | **Primary data** | **Secondary data** |
| Collection purpose | For the problem at hand | For other problems |
| Collection process | Very involved | Rapid and easy |
| Collection cost | High | Relatively low |
| Collection time | Long | Short |
| Suitability | Its suitability is positive | It may or may not suit the object of survey |
| Originality | It is original | It is not original |
| Precautions | No extra precautions required to use the data | It should be used with extra care |

**Limitations of secondary data**

1. Since secondary data is collected for ‘some other purpose, its usefulness to current problem may be limited in several important ways, including relevancies and accuracy.
2. The objectives, nature and methods used to collect secondary data may not be appropriate to present situation.
3. The secondary data may not be accurate, or they may not be completely current or dependable.

**Criteria for evaluating secondary data**

Before using the secondary data it is important to evaluate them on following factors

1. Specification and methodology used to collect the data
2. Error and accuracy of data of the data
3. The currency
4. The objective – The purpose for which data were collected
5. The nature – content of data
6. The dependability

**Sources of data**

**Primary source – The methods of collecting primary data.**

When data is neither internally available nor exists as a secondary source, then the primary sources of data would be approximate.

The various method of collection of primary data are as follows

1. Direct personal investigation
   * Interview
   * Observation
2. Indirect or oral investigation
3. Information from local agents and correspondents
4. Mailed questionnaires and schedules
5. Through enumerations

**Secondary source – The methods of collecting secondary data**

1. Published Statistics
   1. Official publications of Central Government

***Ex***: Central Statistical Organisation (CSO) – Ministry of planning

* + National Sample Survey Organisation (NSSO)
  + Office of the Registrar General and Census Committee – GOI
  + Director of Statistics and Economics – Ministry of Agriculture
  + Labour Bureau – Ministry of Labour etc.

1. Publications of Semi-government organization

***Ex***:

* + The institute of foreign trade, New Delhi
  + The institute of economic growth, New Delhi.

1. Publication of research institutes

***Ex***:

* + Indian Statistical Institute
  + Indian Agriculture Statistical Institute
  + NCRET Publications
  + Indian Standards Institute etc.

1. Publication of Business and Financial Institutions

***Ex***:

* + Trade Association Publications like Sugar factory, Textile mill, Indian chamber of Industry and Commerce.
  + Stock exchange reports, Co-operative society reports etc.

1. News papers and periodicals

***Ex***:

* + The Financial Express, Eastern Economics, Economic Times, Indian Finance, etc.

1. Reports of various committees and commissions

***Ex***:

* + Kothari commission report on education
  + Pay commission reports
  + Land perform committee reports etc.

1. Unpublished statistics
   * Internal and administrative data like Periodical Loss, Profit, Sales, Production Rate, Balance Sheet, Labour Turnover, Budges, etc.

**Classification and Tabulation**

The data collected for the purpose of a statistical inquiry sometimes consists of a few fairly simple figures which can be easily understood without any special treatment. But more often there is an overwhelming mass of raw data without any structure. Thus, unwidely, unorganized and shapeless mass of collected is not capable of being rapidly or easily associated or interpreted. Unorganized data are not fit for further analysis and interpretation. In order to make the data simple and easily understandable the first task is not condense and simplify them in such a way that irrelevant datas are removed and their significant features are stand out prominently. The procedure adopted for this purpose is known as method of classification and tabulation. Classification helps proper tabulation.

“Classified and arranged facts speak themselves; unarranged, unorganised they are dead as mutton”.

- Prof. J.R. Hicks

* **Meaning of Classification**

Classification is a process of arranging things or data in groups or classes according to their resemblances and affinities and gives expressions to the unity of attributes that may subsist among a diversity of individuals.

* **Definition of Classification**

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

The process of grouping large number of individual facts and observations on the basis of similarity among the items, is called classification.

- Stockton & Clark

**Characteristics of classification**

* 1. Classification performs homogeneous grouping of data
  2. It brings out points of similarity and dissimilating
  3. The classification may be either real or imaginary
  4. Classification is flexible to accommodate adjustments

**Objectives / purposes of classifications**

1. To simplify and condense the large data
2. To present the facts to easily in understandable form
3. To allow comparisons
4. To help to draw valid inferences
5. To relate the variables among the data
6. To help further analysis
7. To eliminate unwanted data
8. To prepare tabulation

**Guiding principles (rules) of classifications**

Following are the general guiding principles for good classifications

* 1. ***Exhaustive***: Classification should be exhaustive. Each and every item in data must belong to one of class. Introduction of residual class (i.e. either, miscellaneous etc.) should be avoided.
  2. ***Mutually exclusive***: Each item should be placed at only one class
  3. ***Suitability***: The classification should confirm to object of inquiry.
  4. ***Stability***: Only one principle must be maintained throughout the classification and analysis.
  5. ***Homogeneity***: The items included in each class must be homogeneous.
  6. ***Flexibility***: A good classification should be flexible enough to accommodate new situation or changed situations.

**Modes / Types of Classification**

Modes / Types of classification refers to the class categories into which the data could be sorted out and tabulated. These categories depend on the nature of data and purpose for which data is being sought.

**Important types of classification**

1. Geographical (i.e. on the basis of area or region wise)
2. Chronological (On the basis of Temporal / Historical, i.e. with respect to time)
3. Qualitative (on the basis of character / attributes)
4. Numerical, quantitative (on the basis of magnitude)
5. ***Geographical Classification***

In geographical classification, the classification is based on the geographical regions.

***Ex***: Sales of the company (In Million Rupees) (region – wise)

|  |  |
| --- | --- |
| **Region** | **Sales** |
| North | 285 |
| South | 300 |
| East | 185 |
| West | 235 |

1. ***Chronological Classification***

If the statistical data are classified according to the time of its occurrence, the type of classification is called chronological classification.

Sales reported by a departmental store

|  |  |
| --- | --- |
| **Month** | **Sales**  **(Rs.) in lakhs** |
| January | 22 |
| February | 26 |
| March | 32 |
| April | 25 |
| May | 27 |
| June | 29 |
| July | 30 |
| August | 30 |

1. ***Qualitative Classification***

In qualitative classifications, the data are classified according to the presence or absence of attributes in given units. Thus, the classification is based on some quality characteristics / attributes.

**Ex**: Sex, Literacy, Education, Class grade etc.

Further, it may be classified as

a) Simple classification b) Manifold classification

1. *Simple classification*: If the classification is done into only two classes then classification is known as simple classification.

***Ex:*** a) Population in to Male / Female

b) Population into Educated / Uneducated

1. *Manifold classification*: In this classification, the classification is based on more than one attribute at a time.

***Ex***:

**Population**

**Smokers**

**Non-smokers**

**Illiterate**

**Literate**

**Male**

**Female**

**Male**

**Female**

**Literate**

**Illiterate**

**Male**

**Female**

**Male**

**Female**

1. ***Quantitative Classification***: In Quantitative classification, the classification is based on quantitative measurements of some characteristics, such as age, marks, income, production, sales etc. The quantitative phenomenon under study is known as variable and hence this classification is also called as classification by variable.

***Ex***:

For a 50 marks test, Marks obtained by students as classified as follows

|  |  |
| --- | --- |
| **Marks** | **No. of students** |
| 0 – 10 | 5 |
| 10 – 20 | 7 |
| 20 – 30 | 10 |
| 30 – 40 | 25 |
| 40 – 50 | 3 |
| **Total Students = 50** | |

In this classification marks obtained by students is variable and number of students in each class represents the frequency.

**Meaning and Definition of Tabulation**

Tabulation may be defined as systematic arrangement of data is column and rows. It is designed to simplify presentation of data for the purpose of analysis and statistical inferences.

**Major Objectives of Tabulation**

1. To simplify the complex data
2. To facilitate comparison
3. To economise the space
4. To draw valid inference / conclusions
5. To help for further analysis

**Differences between Classification and Tabulation**

1. First data are classified and presented in tables; classification is the basis for tabulation.
2. Tabulation is a mechanical function of classification because is tabulation classified data are placed in row and columns.
3. Classification is a process of statistical analysis while tabulation is a process of presenting data is suitable structure.

**Classification of tables**

Classification is done based on

1. Coverage (Simple and complex table)
2. Objective / purpose (General purpose / Reference table / Special table or summary table)
3. Nature of inquiry (primary and divided table).

***Ex:***

1. *Simple table*: Data are classified based on only one characteristic

Distribution of marks

|  |  |
| --- | --- |
| **Class Marks** | **No. of students** |
| 30 – 40 | 20 |
| 40 – 50 | 20 |
| 50 – 60 | 10 |
| **Total** | **50** |

1. *Two-way table*: Classification is based on two characteristics

|  |  |  |  |
| --- | --- | --- | --- |
| **Class Marks** | **No. of students** | | |
| **Boys** | **Girls** | **Total** |
| 30 – 40 | 10 | 10 | 20 |
| 40 – 50 | 15 | 5 | 20 |
| 50 – 60 | 3 | 7 | 10 |
| **Total** | **28** | **22** | **50** |

**Frequency Distribution**

Frequency distribution is a table used to organize the data. The left column (called classes or groups) includes numerical intervals on a variable under study. The right column contains the list of frequencies, or number of occurrences of each class/group. Intervals are normally of equal size covering the sample observations range.

It is simply a table in which the gathered data are grouped into classes and the number of occurrences which fall in each class is recorded.

* **Definition**

A frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups with their corresponding frequencies.

- Croxton and Cowden

A frequency distribution can be classified as

1. Series of individual observation
2. Discrete frequency distribution
3. Continuous frequency distribution
4. ***Series of individual observation***

Series of individual observation is a series where the items are listed one after the each observation. For statistical calculations, these observations could be arranged is either ascending or descending order. This is called as array.

***Ex***:

|  |  |
| --- | --- |
| **Roll No.** | **Marks obtained in statistics paper** |
| 1 | 83 |
| 2 | 80 |
| 3 | 75 |
| 4 | 92 |
| 5 | 65 |

The above data list is a raw data. The presentation of data in above form doesn’t reveal any information. If the data is arranged in ascending / descending in the order of their magnitude, which gives better presentation then, it is called arraying of data.

**Discrete (ungrouped) Frequency Distribution**

If the data series are presented in such a way that indicating its exact measurement of units, then it is called as discrete frequency distribution. Discrete variable is one where the variates differ from each other by definite amounts.

***Ex***: Assume that a survey has been made to know number of post-graduates in 10 families at random, the resulted raw data could be as follows.

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

This data can be classified into an ungrouped frequency distribution. The number of post-graduates becomes variable (x) for which we can list the frequency of occurrence (f) in a tabular from as follows;

|  |  |
| --- | --- |
| **Number of post graduates (x)** | **Frequency (f)** |
| 0 | 2 |
| 1 | 2 |
| 2 | 4 |
| 3 | 1 |
| 4 | 1 |

The above example shows a discrete frequency distribution, where the variables has discrete numerical values.

**Continuous frequency distribution (grouped frequency distribution)**

Continuous data series is one where the measurements are only approximations and are expressed in class intervals within certain limits. In continuous frequency distribution the class interval theoretically continuous from the starting of the frequency distribution till the end without break. According to Boddington ‘the variable which can take very intermediate value between the smallest and largest value in the distribution is a continuous frequency distribution.

***Ex***:

Marks obtained by 20 students in exam for 50 marks are as given below – convert the data into continuous frequency distribution form.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 18 | 23 | 28 | 29 | 44 | 28 | 48 | 33 | 32 | 43 |
| 24 | 29 | 32 | 39 | 49 | 42 | 27 | 33 | 28 | 29 |

By grouping the marks into class interval of 10 following frequency distribution table can be formed.

|  |  |
| --- | --- |
| **Marks** | **No. of students** |
| 0 - 5 | 0 |
| 5 – 10 | 0 |
| 10 – 15 | 0 |
| 15 – 20 | 1 |
| 20 – 25 | 2 |
| 25 – 30 | 7 |
| 30 – 35 | 4 |
| 35 – 40 | 1 |
| 40 – 45 | 3 |
| 45 – 50 | 2 |

**Technical terms used in formulation frequency distribution**

1. ***Class limits:***

The class limits are the smallest and largest values in the class.

***Ex***: 0 – 10, in this class, the lowest value is zero and highest value is 10. the two boundaries of the class are called upper and lower limits of the class. Class limit is also called as class boundaries.

1. ***Class intervals***

The difference between upper and lower limit of class is known as class interval.

***Ex***:

In the class 0 – 10, the class interval is (10 – 0) = 10.

The formula to find class interval is gives on below



L = Largest value S = Smallest value R = the no. of classes

***Ex***: If the marks of 60 students in a class varies between 40 and 100 and if we want to form 6 classes, the class interval would be

i= (L-S)/K =  =  = 10 L = 100

S = 40

R = 6

Therefore, class intervals would be 40 – 50, 50 – 60, 60 – 70, 70 – 80, 80 – 90 and 90 – 100.

* **Methods of forming class-interval**

1. ***Exclusive method (overlapping)***

In this method, the upper limits of one class-interval is the lower limit of next class. This methods makes continuity of data.

***Ex***:

|  |  |
| --- | --- |
| **Marks** | **No. of students** |
| 20 – 30 | 5 |
| 30 – 40 | 15 |
| 40 – 50 | 25 |

A student whose mark is between 20 to 29.9 will be included in the 20 – 30 class.

Better way of expressing is

|  |  |
| --- | --- |
| **Marks** | **No. of students** |
| 20 to less than 30  (More than 20 but less than 30) | 5 |
| 30 to less than 40 | 15 |
| 40 to less than 50 | 25 |
| **Total Students** | **50** |

1. ***Inclusive method (non-overlaping)***

***Ex***:

|  |  |
| --- | --- |
| **Marks** | **No. of students** |
| 20 – 29 | 5 |
| 30 – 39 | 15 |
| 40 – 49 | 25 |

A student whose mark is 29 is included in 20 – 29 class interval and a student whose mark in 39 is included in 30 – 39 class interval.

* **Class Frequency**

The number of observations falling within class-interval is called its class frequency.

***Ex***: The class frequency 90 – 100 is 5, represents that there are 5 students scored between 90 and 100. If we add all the frequencies of individual classes, the total frequency represents total number of items studied.

* **Magnitude of class interval**

The magnitude of class interval depends on range and number of classes. The range is the difference between the highest and smallest values is the data series. A class interval is generally in the multiples of 5, 10, 15 and 20.

Sturges formula to find number of classes is given below

K = 1 + 3.322 log N.

K = No. of class

log N = Logarithm of total no. of observations

***Ex***: If total number of observations are 100, then number of classes could be

K = 1 + 3.322 log 100

K = 1 + 3.322 x 2

K = 1 + 6.644

K = 7.644 = 8 (Rounded off)

**NOTE**: Under this formula number of class can’t be less than 4 and not greater than 20.

* **Class mid point or class marks**

The mid value or central value of the class interval is called mid point.

Mid point of a class = 

* **Sturges formula to find size of class interval**

Size of class interval (h) = 

***Ex***: In a 5 group of worker, highest wage is Rs. 250 and lowest wage is 100 per day. Find the size of interval.

h =  = = 55.57≅ 56

**Constructing a frequency distribution**

The following guidelines may be considered for the construction of frequency distribution.

1. The classes should be clearly defined and each observations must belong to one and to only one class interval. Interval classes must be inclusive and non-overlapping.
2. The number of classes should be neither too large nor too small.

Too small classes result greater interval width with loss of accuracy. Too many class interval result is complexity.

1. All interval should be of the same width. This is preferred for easy computations.

The width of interval = 

1. Open end classes should be avoided since creates difficulty in analysis and interpretation.
2. Intervals would be continuous throughout the distribution. This is important for continuous distribution.
3. The lower limits of the class intervals should be simple multiples of the interval.

***Ex***: A simple of 30 persons weight of a particular class students are as follows. Construct a frequency distribution for the given data.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 62 | 58 | 58 | 52 | 48 | 53 | 54 | 63 | 69 | 63 |
| 57 | 56 | 46 | 48 | 53 | 56 | 57 | 59 | 58 | 53 |
| 52 | 56 | 57 | 52 | 52 | 53 | 54 | 58 | 61 | 63 |

* **Steps of construction**

***Step 1***

Find the range of data (H) Highest value = 70

(L) Lowest value = 46

Range = H – L = 69 – 46 = 23

***Step 2***

Find the number of class intervals.

Sturges formula

K = 1 + 3.322 log N.

K = 1 + 3.222 log 30

K = 5.90 Say K = 6

**∴ No. of classes = 6**

***Step 3***

Width of class interval

Width of class interval =  = 

***Step 4***

Conclusions all frequencies belong to each class interval and assign this total frequency to corresponding class intervals as follows.

|  |  |  |
| --- | --- | --- |
| **Class interval** | **Tally bars** | **Frequency** |
| 46 – 50 | | | | | 3 |
| 50 – 54 | | | | | | | | | 8 |
| 54 – 58 | | | | | | | | | 8 |
| 58 – 62 | | | | | | | 6 |
| 62 – 66 | | | | | | 4 |
| 66 – 70 | | | 1 |

**Cumulative frequency distribution**

Cumulative frequency distribution indicating directly the number of units that lie above or below the specified values of the class intervals. When the interest of the investigator is on number of cases below the specified value, then the specified value represents the upper limit of the class interval. It is known as ‘less than’ cumulative frequency distribution. When the interest is lies in finding the number of cases above specified value then this value is taken as lower limit of the specified class interval. Then, it is known as ‘more than’ cumulative frequency distribution.

The cumulative frequency simply means that summing up the consecutive frequency.

***Ex***:

|  |  |  |
| --- | --- | --- |
| **Marks** | **No. of students** | **‘Less than’ cumulative frequency** |
| 0 – 10 | 5 | 5 |
| 10 – 20 | 3 | 8 |
| 20 – 30 | 10 | 18 |
| 30 – 40 | 20 | 38 |
| 40 – 50 | 12 | 50 |

In the above ‘less than’ cumulative frequency distribution, there are 5 students less than 10, 3 less than 20 and 10 less than 30 and so on.

Similarly, following table shows ‘greater than’ cumulative frequency distribution.

***Ex***:

|  |  |  |
| --- | --- | --- |
| **Marks** | **No. of students** | **‘More than’ cumulative frequency** |
| 0 – 10 | 5 | 50 |
| 10 – 20 | 3 | 45 |
| 20 – 30 | 10 | 42 |
| 30 – 40 | 20 | 32 |
| 40 – 50 | 12 | 12 |

In the above ‘greater than’ cumulative frequency distribution, 50 students are scored more than 0, 45 more than 10, 42 more than 20 and so on.