Frequency: Frequency, also called class frequency refers to the number of observations falling within a particular class. It is the number of measurements or counts in a category or class.

Frequency distribution: In statistics, frequency distribution is a graph or a data set organized to show the frequency of occurrence of each possible outcome of a repeatable event observed many times.

→ A frequency distribution in statistics is a representation that displays the number of observations within a given interval.

Frequency distribution can be constructed for both categorical and numerical data.

When numerical data is grouped and organized in a frequency distribution results, it is called grouped frequency distribution.

For ungrouped data, we refer to as ungrouped frequency distribution.

constructing of Frequency table using categorical data (qualitative data)

→ The constructing of a frequency table for categorical data with a single categorical variable consists essentially of the following steps:

(a) Choose the category into which the data are to be grouped.

(b) Sort or tally the data into appropriate categories.

(c) count the number of items or measurements falling in each category.

(d) Display the results in a table.

(e) The resulting table represents the desired frequency distribution.

2.2 Example: A market researcher conducted an inventory of 25 firms and categorized them as 'large', 'medium' and 'small' depending on the investment, floor space and number of employees. The categories of the 25 listed firms were as follows:

small	large	small	medium	large
medium	large	small	large	medium
large	small	large	medium	medium
medium	large	large	medium	medium
medium	small	small	medium	medium

Present the data in a frequency distribution.

Solution:

Here, the data pertain to three distinct categories of the firm and they are large, medium and small.

count the tallies for each category following the table:

Firm size	Tally	Count
Large	𝍸 III	8
Medium	𝍸 𝍸 I	11
Small	𝍸 I	6

Frequency distribution of the firm size:

Firm size	Number of Firms	Percent
large	8	32.0
Medium	11	44.0
Small	6	24.0
Total	25	100.0

$(\frac{8}{25} \times 100)$

Here, the number firms (8, 11, 6) represent the class or frequency for the categories large, medium, small respectively. The count 25 rep is the total frequency.

Since, the data is grouped, it is known as categorical distribution.

Example-2.3: A market research team conducted a servey among 350 boys and 250 girls of about the same age on their preference of cold drink available in the market.

	Types of Drink				
Sex	coca-cola	Sprite	Mountain dew	7-up	Total
Boys	95	90	85	80	350
Girls	65	50	80	55	250
Total	160	140	165	135	600

(a) of the total children, how many of the girls prefer sprite?

⇒ Here, 50 girls out of 600 like sprite.

∴ percentage $= \frac{50}{600} \times 100\% = 8.3\%$

(b) How many of the children like mountain dew?

⇒ percentage $= \frac{165}{600} \times 100\% = 27.5\%$

(c) Among the total children, what is the percentage of boys who prefer 7-up?

⇒ percentage $= \frac{80}{600} \times 100\% = 13.3\%$

(d) Among the total children, how many do prefer sprites?

⇒ percentage $= \frac{140}{600} \times 100\% = 23.3\%$

(e) of the total boys, what percentage of them like mountain dew?

⇒ percentage $= \frac{85}{350} \times 100\% = 24.3\%$

(f) of those who prefer coca-cola, what is the percentage of boys?

⇒ percentage $= \frac{95}{160} \times 100\% = 59.4\%$

(9) construct the marginal distribution for boys and girls.

Drink type	No. of children	Percent
coca-cola	160	26.7
sprite	140	23.3
Mountain dew	165	27.5
7-up	135	22.5
Total	600	100

constructing of frequency distribution using numerical data

→ The process of constructing a frequency distribution with numerical or quantitative data is very similar to those for qualitative data, except that now the data have to be grouped into classes of appropriate intervals. The simplest device in doing so is to form an array first.

An array is an ordering of values of the variables in order of their magnitude, usually in ascending order; from smallest to the largest.

4.

ungrouped data

50	63	70	75	84
51	65	71	75	85
54	65	72	76	86
56	66	72	77	87
56	67	72	79	88
57	68	73	80	88
59	68	73	81	89
60	69	74	82	93
61	69	74	82	93
62	70	74	83	97

ungrouped frequency distribution:

Wage	Frequency
50	1
54	1
56	1
56	2
57	1
⋮	⋮
65	2
⋮	⋮
68	2
69	2
70	2
⋮	⋮
72	3
73	2
74	3
75	2
⋮	⋮
⋮	⋮
97	1

Various terms of frequency distributions for numerical data:

class: A class is an interval containing observations, each observation being classified into one and only one class.

Frequency: The number of observations or values falling into each group or class is called class frequency. The frequency of a class thus shows how many times a particular value of observation is repeated in that class.

class limits: For numerical data, the frequencies of a particular class are bounded by two values. The smaller value of the class is known as the lower class limit while the larger value is known as the upper class limit.

class boundary:

class interval: The width (w) or length of the class, formed by two boundary values is known as the class interval or class width. A class interval represents the spread between the class boundaries.

class mark: The class mark is the value that lies in the middle of the class and is obtained by averaging the two class boundaries. The class is also referred to as clas-mid-point or mid-value of the class.

open interval: An open interval is an interval with one of it's limits (in either side) indeterminate. Thus an age of a person recorded as less than 45 years (<45) constitute an open interval.

Steps in constructing grouped frequency distribution

(a) Decide on number of classes and the class width in which the observations are to be grouped.

(b) Assign the observations to the appropriately chosen classes. This called tallying.

(c) count the number of observation's falling in each class. These numbers are the frequencies.

(d) Display the results obtained in the above three steps in a table or a chart.

(e) The resulting table is our deserved frequency distribution.

If the smallest value(S) and the largest value(L) in a data set are known, then as a rule of thumb,

the range, $R = L - S$

class width, (W) & Approximate number of classes (K)

$\therefore K = \frac{L-S}{W} = \frac{R}{W}$

Determine, $K = 1 + 3.322 \log N$

$$W = \frac{R}{1 + 3.322 \log_{10} N}$$

This approach has been proposed by Sturge: Another empirical rule suggested by Sturge to determine the number of classes is the "2 to the K rule". This rule suggests that the number of classes should be the smallest whole number K that makes the quantity 2^K greater than the total number of observations (N) in the data set, that is $2^K \geq N$. Suppose that, $N = 50$ observations. Then since $2^5 = 32$ which is smaller than N and $2^6 = 64$, which is greater than N. The Sturge's rules also dictates us to choose 6 classes, that is $K = 6$

Constructing frequency distribution using continuous data

Ex-2.5: construct a suitable frequency distribution with class width of appropriate size:

The ages of the 50 workers

25	33	37	42	45
28	34	37	42	46
29	35	37	42	46
30	35	38	43	46
31	35	38	43	46
32	36	38	43	47
32	36	39	44	50
32	36	40	44	51
33	36	41	44	52
33	37	42	45	54

Sol: From giving data,

$N = 50$

number of classes, $K = 1 + 3.322 \log 50 = 6.64$

which is approximately 6. $\therefore K = 6$

$\therefore$ Range, R = upper limit (L) − lower limit (S)

$= (54 - 25) = 29$

$\therefore$ class width, $W = \frac{R}{K} = \frac{29}{6} = 4.83$

which is approximately 5 $\therefore W = 5$

Table: Age Distribution

Age in Years	Tally	Number of Workers
25-29	III	3
30-34	𝍸 IIII	9
35-4039	𝍸 𝍸 𝍸 I	15
40-44	𝍸 𝍸 II	12
45-49	𝍸 II	7
50-54	IIII	4
		Total: 50

Other forms of frequency distribution:

1. Percentage frequency distribution
2. Relative frequency distribution
3. Cumulative distribution

Percentage frequency distribution: A percentage frequency distribution is formed by dividing the number of classes attribute to a category by the total number of cases and multiplying the resulting value by 100. Thus f_i is the frequency of the i-th class of a frequency distribution and N the total frequency then the percentange of i-th class is,

$$P_i = \frac{f_i}{N} \times 100$$

∴ The total frequency in a percent distribution that is $= \sum\left(\frac{f_i}{N}\right) \times 100 = 100$

Relative frequency distribution: Instead of presenting the frequencies in absolute figures, it is sometimes convenient to express the frequencies in related terms. The resulting distribution is then called relative frequency distribution. The relative frequency is simply the fraction or proportion of the total number of items belonging to the class or category. For a data set having a total number of observation, N. The related frequency of i-th class is,

Relative frequency of i-th class $= \frac{f_i}{N}$

The total relative frequency in such a distribution is, $\sum\left(\frac{f_i}{N}\right) = 1.0$

[The relative frequencies are essentially proportions when multiplied by 100, result in percentage frequencies and hence the percentage distribution.]

Cumulative frequency distribution: In statistics, a cumulative frequency distribution is defined as the total of frequencies, that are distributed over different class intervals. There are two type of cumulative frequency distribution and they are:

1. less than type
2. more than type

less than type frequency distribution: the less than cumulative frequency distribution is obtained by adding successively the frequencies of all the previous classes along with the class.

More than type frequency distribution: The more than cumulative frequency distribution is obtained by determining the total frequencies starting from the highest class to the lowest class.

Table: Frequency distribution

Class boundaries	Absolute frequency	Percentage frequency	Relative frequency	less than type C.F	% less than type C.F	More than type C.F	% more than type C.F
49.5-57.5	6	12.0	0.12	6	12.0	50	100.0
57.5-65.5	7	14.0	0.14	13	26.0	44	88.0
65.5-73.5	14	28.0	0.28	27	54.0	37	74.0
73.5-81.5	10	20.0	0.20	37	74.0	23	46.0
81.5-89.5	10	20.0	0.20	47	94.0	13	26.0
89.5-97.5	3	6.0	0.06	50	100	3	6.0
Total:	50	100	1.0				

percentange frequency for 3rd class $= \frac{f_3}{N} \times 100\% = \frac{14}{50} \times 100\% = 28\%$

relative frequency $= \frac{f_3}{N} = \frac{14}{50} = 0.28$

Graphical Presentation of Data

→ In addition to presenting a frequency distribution in tabular form, one can present the same through some visual aids. This refers to graphs and diagrams or chart.
As we aware that a frequency distribution can be constructed either from categorical (qualitative) and numerical (quantitive) data, the graphs and diagrams to be constructed will also differ accordingly: categorical or numerical.

Presenting of categorical Data:

→ Bar diagram, stacked bar diagram
→ Pie diagram

Bar diagram: Bar diagram also called bar charts are commonly used to describe categorical data. A bar diagram is a form of presentation in which the frequencies against the categories are represented by rectangles seperated usually along the horizontal axis and drawns as bars of convenient widths. The width of these bars have significance but are taken to make the chart look attractive.
We represent the data by two different bar diagrams. These are,

(a) Vertical bars
(b) Horizontal bars

Ex-2.11: The accompanying table shows the stock position of finished goods in Metric tons as of June 2004 of the Bangladesh chemical industry (BCIC). Represent the data by suitable diagram.

Finished Goods	Quantity (in Metric Ton)
TSP	8916
SSP	18455
Paper	2660
Cement	7048
Sanitary ware	1620
Insulator	3520
Tiles	17335
Total	54928

Sol:

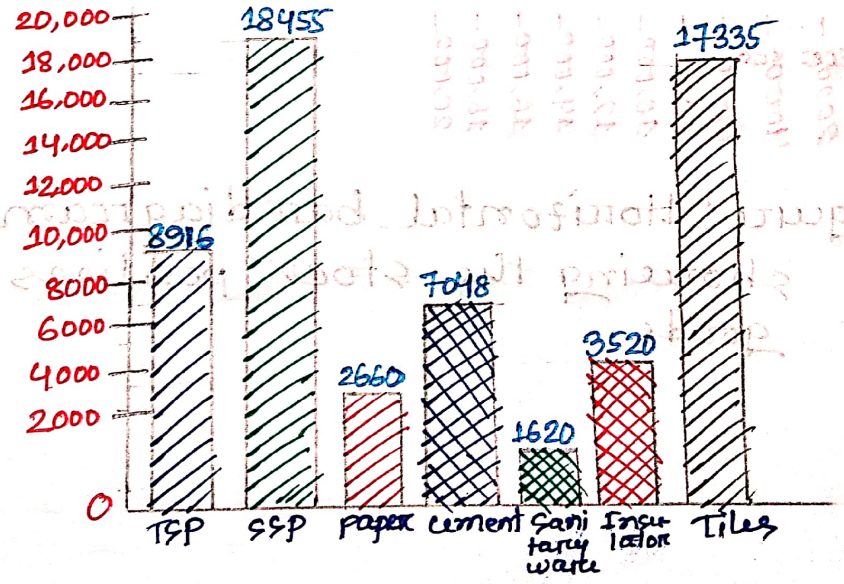

Figure: Vertical bar chart showing the stock positions of goods.

Alternative:

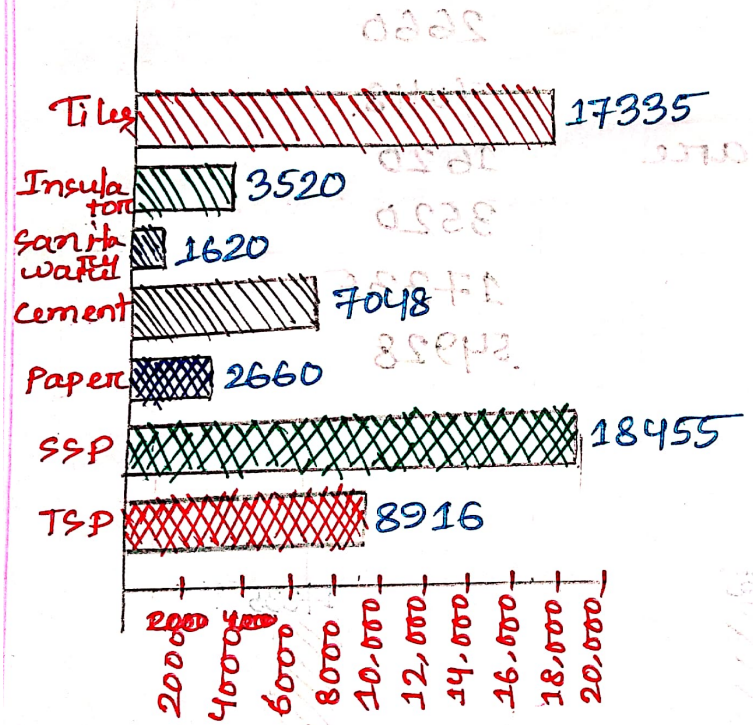

Figure: Horizontal bar diagram showing the stock positions of goods