machine does not only tabulate the data at high speed, but is also capable of printing totals and sub-totals, and may be so set as to reproduce any part of the data contained in the cards at speeds ranging from 5,000 to 12,000 cards per hour.

EXERCISES

- 1. Examine whether the following variables are discrete or continuous—Age of a person, Size of land holding, Size of family, Temperature, Monthly sales in a shop.

 [C.U., M.Com. '69]
- 2. Explain the terms 'classification' and 'tabulation' and point out their importance in a statistical investigation. What precautions would you take in tabulating statistical data? [1.C.W.A., Jan. '70]
- 3. State the various reasons why errors are made in the interpretation of statistical data, and indicate how to guard against such errors. [C.A., May '70 & '72]
- 4. What do you mean by an error in statistics? Classify the different types of errors in statistics and discuss their effect on statistical inference.

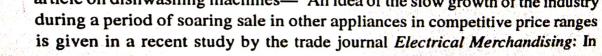
[C.A., May '76]

5. Draw up a table to show the number of wholly unemployed, temporarily stopped and the total unemployed, each class being divided into males and females for the following industries: Fishing, Coal mining, Iron-ore mining, Engineering and Ship building.



- 6. Draw up a blank table in which could be shown the number of persons employed in six industries, on two different dates, distinguishing males from females and among the latter, singles, married and widows.
- 7. Prepare a blank table showing the distribution of population according to sex and four religions in five age groups in seven different cities.
- 8. Draw up a blank table showing the Exports and Imports during the years 1960, 1961, 1962, 1963 and 1964 relating to the ports Mumbai, Kolkata, Madras and other ports. The table should provide for the values and the balance of trade and the totals for each year.
- 9. You are given data on exports (both quantity and value) of Indian jute to U.K., U.S.A., U.S.S.R., Japan and Canada for 5 consecutive years. Suggest a suitable tabular representation by drawing a blank table. [B.U., B.A. (Econ) '72]
- 10. Draft a blank table to show the distribution of personnel in a manufacturing concern according to (a) Sex: male and female; (b) Three grades of salary: below Rs 300, Rs 300-Rs 500, Rs 500 and over; (c) Two periods: 1940 and 1950; (d) Three age groups: below 25, 25 and under 40, 40 and over.
- [C.A., May '63]

 11. Tabulate the following information which appeared as part of a newspaper article on dishwashing machines—"An idea of the slow growth of the industry during a period of soaring sale in other appliances in competitive price ranges



1947, the survey showed, 120,000 motor-driven dishwashers were sold at an average price of \$ 250. In 1948, sales rose to 225,000 at an average price of \$ 275. In 1949, with the average price the same as the year before, the sales dropped to 160,000. In 1950, they jumped to 230,000 at an average price of \$ 290 and in 1951 they continued to climb to 260,000 units even though the average price went up to \$ 300. In 1952, sales dropped to 175,000 units and in 1953 they rose to only 180,000 despite the fact that average price remained at \$ 300."

12. Present the following information in a tabular form and suggest a suitable title:

'The production of 10.95 lakh tons of rice in Maharashtra in 1962–1963 was the lowest in the period since 1955–56. In 1963–64, however, it has shown a spectacular recovery and reached the level of 15.14 lakh tons. During 1963–64, wheat and bajri output decreased. The production of bajri which was 5.50 lakh tons in 1962–63 declined to 4.51 lakh tons in 1963–64. The production of wheat also decreased from 4.63 lakh tons in 1962–63 to 3.43 lakh tons in 1963–64. The area under pulses has shown a decreasing trend and the production was less by 22,000 tons in 1963–64 than the production of 8.89 lakh tons in 1962–63.'

[I.C.W.A., Jan. '70]

13. An investigation conducted by the Education Department in a public library revealed the following facts. You are required to tabulate the information as neatly and clearly as you can:





"In 1950 the total number of the readers was 46,000 and they borrowed some 16,000 volumes. In 1960 the number of books borrowed increased by 4,000 and the borrowers by 50%. The classification was on the basis of three sections: Literature, Fiction and Illustrated News. There were 10,000 and 30,000 readers in the sections Literature and Fiction respectively in the year 1950. In the same year 2,000 and 10,000 books were lent in the sections Illustrated News and Fiction respectively. Marked changes were seen in 1960. There were 7,000 and 42,000 readers in the Literature and Fiction sections respectively. So also 4,000 and 13,000 books were lent in the sections Illustrated News and Fiction respectively."

14. Tabulate the following information which has been taken from *The Times Review of Industry*, 1958:

'In 1958 France exported 1,66,000 cars as against 1,07,000 in 1957 of which 64,000 went to Europe as compared with 46,000 in the same period in 1957. Some 34,000 went to oversea French territories as compared with 26,000 and 68,000 elsewhere as against 35,000. The increase to countries other than Europe is 40 per cent. The proportion of French production going to foreign markets was 27 per cent, last year. Renault's sales outside France amounted to 37 per cent of total production, while their total foreign sales of cars was 92,000.'

[C.U., B.A. (Econ) '69]

15. The total population of Hong Kong aged 15 and over rose from 1853 thousand in 1961 to 2528 thousand in 1971. The male population rose from 945 thousand to 1280 thousand and the female population from 908 thousand to 1248.



thousand. The distribution of population (above 15 years) by conjugal status was as follows in 1961: 504 thousand persons were never married, 1202 thousand were currently married, 135 thousand widowed and only 12 thousand divorced or separated. The corresponding figures (in thousand) were 914, 1481, 125 and 8 in 1971. Among males aged 15 or more in 1961, 334 thousand were never married, 588 thousand currently married, 17 thousand widowed and 6 thousand divorced or separated; the corresponding figures for 1971 were 546, 710, 20 and 4 (in thousand).

Present the data in a tabular form with appropriate headings and sub-headings.

[C.U., B.A.(Econ) '78]

16. The following is an excerpt from the White Paper on the general budget for 1957-58, appearing in the publication "Budget for 1957-58" of the Ministry of Finance, Govt. of India:

"...The value of imports increased from Rs 418 crores in April-November, 1955, to Rs 535 crores in April-November, 1956. Of this increase of Rs 117 crores, Rs 96 crores was accounted for by the increase in the imports of machinery, iron and steel and other metals. Imports of machinery increased from Rs 73.5 crores in April-November, 1955 to Rs 105.8 crores in April-November, 1956; of iron and steel from Rs 34.3 crores to Rs 88 crores and of other metals from Rs 16.4 crores to Rs 26.2 crores.

The small decline in exports (from Rs 388.4 crores in April-November, 1955, to 378 crores in April-November, 1956) is mostly accounted for by



G Statistical Methods (Combined) - ×	Ť		
c=frontcover#v=onepage&q&f=false			
	Q		
\dd to my library ▼ Write review			Page 24 🔻

the decline in exports of oil (from Rs 26.2 crores to Rs 13.4 crores) and of cotton (from Rs 24.6 crores to Rs 10.9 crores) which is largely explained by the poor crop of oil seeds and cotton in 1955–56. Exports of manganese ore also declined from Rs 10.6 crores in April-November, 1955 to Rs 6.8 crores in April-November, 1956; of cotton textiles from Rs 42.2 crores to Rs 40.1 crores and of jute manufactures from Rs 83.3 crores to Rs 79.5 crores."

Represent the information contained in the above excerpt in a suitable tabular form.

[B.U., B.A.(Econ) '68]

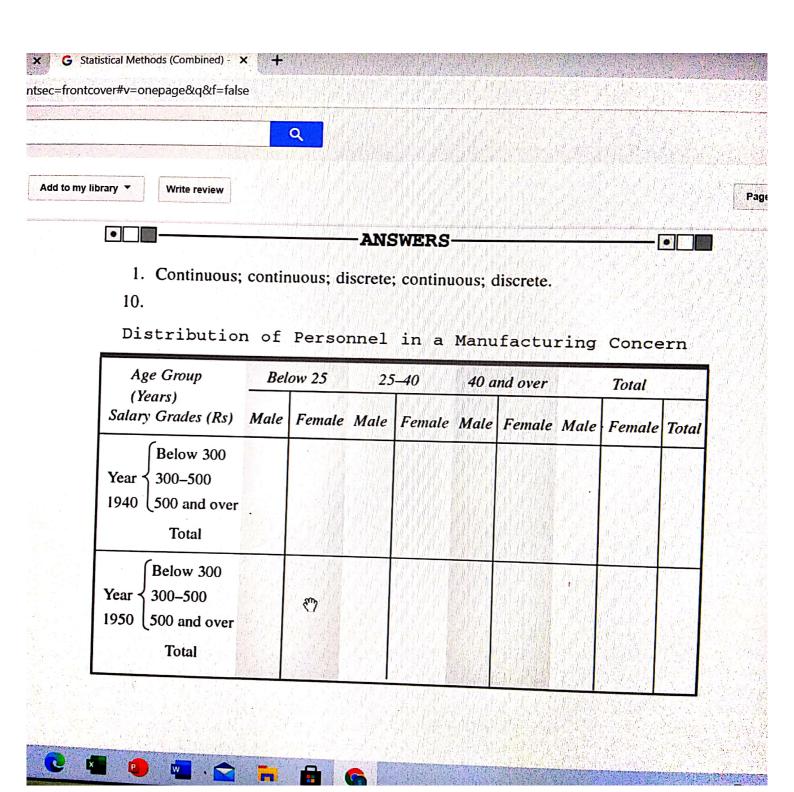
17. Present the following information in a concise tabular form and indicate which type of lamp shows the greatest wastage during manufacture:

'Lamps are rejected at several manufacturing stages for different faults. 12,000 glass tubes are supplied to make 40-watt, 60-watt and 100-watt lamps in the ratio 1:2:3. At the stage I, 10 per cent of the 40-watt, 4 per cent of the 60-watt and 5 per cent of the 100-watt bulbs are broken. At the stage II, about 1 per cent of the remainder of the lamps have broken filaments. At the stage III, 100 100-watt lamps have badly soldered caps, and half as many have croooked caps; twice as many 40-watt and 60-watt lamps have these faults. At the stage IV, about 3 per cent are rejected for bad type-marking and 1 in every 100 are broken in the packing which follows.'

[I.C.W.A., July '71; C.U., B.A.(Econ) '67]

18. Mention different methods of tabulation applied to survey data.

[I.C.W.A., Dec. '73]



Collection of Data

25

11.

Sale of Motor-driven Dishwashers (1947-1953)

Year	Number of Units (1	.947-1953)
	sold ('000)	Average Price
1947	The Property of the second property of the property of the second pr	(\$)
1948	120	250
1949	225	275
1950	160	275
1951	230	290
1952	260	300
1953	175 180	300

Source: Electrical Merchandising

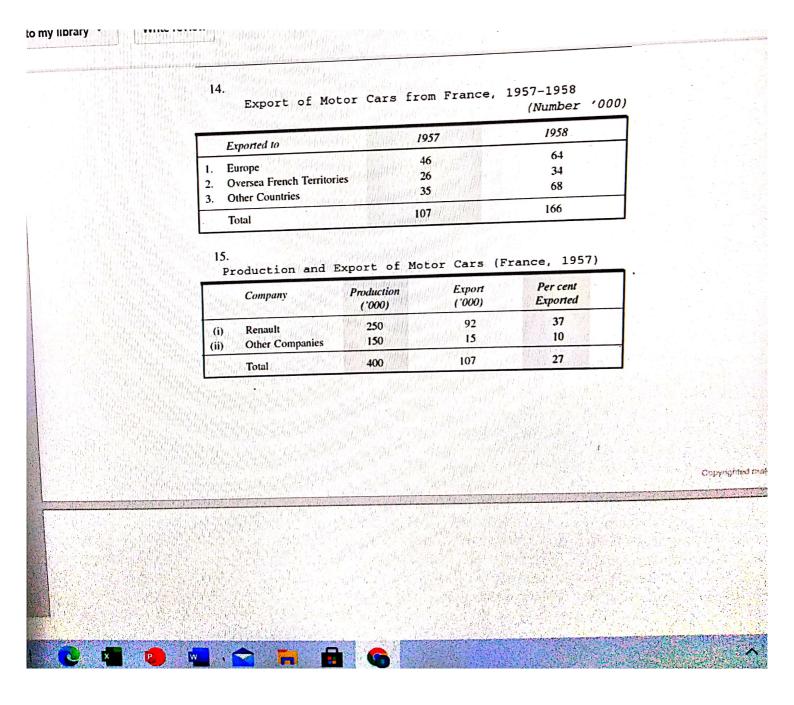
13.

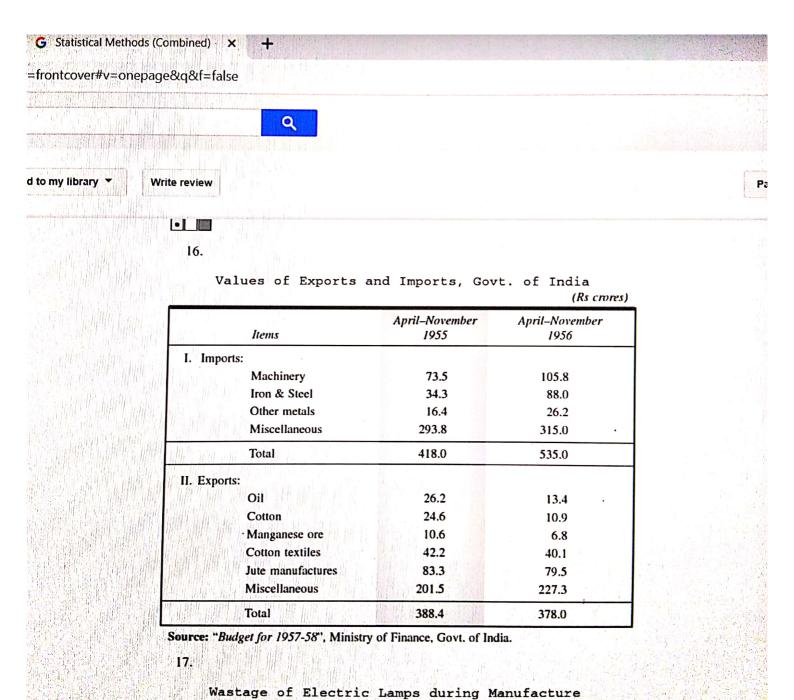
Number of Readers in and Books Borrowed Library (Years from a Public 1950 & 1960)

Sections	Readers ('000)		Books Bor	rowed ('000)
	1950	1960	1950	1960
 Literature Fiction Illustrated News 	10 30 6	7 42 20	4 10	3 13
Total	46	69	16	20

14.







Add to my library ▼

Write review

167	Manganese ore 10.6 Cotton textiles 42.2 Jute manufactures 83.3 Miscellaneous 201.5	6.8 40.1 79.5 227.3	
	Total 388.4	378.0	

Source: "Budget for 1957-58", Ministry of Finance, Govt. of India.

17.

Wastage of Electric Lamps during Manufacture

Patroniai ()		WW. J.	Number Rej	ected at S	tages	Add William	
Power of Lamp (watt)	Total No. of Glass Tubes Supplied	I Broken bulbs	II Broken filament	III Bad cap	IV Bad type marking	Total	Rejections %
40	2,000	200	18	300	59	577	29
60	4,000	160	38	300	140	638	16
100	6,000	300	57	150	220	727	12
Total	12,000	600	113	750	419	1,942	16

⁴⁰⁻watt lamps have the greatest wastage, viz. 29%.

```
(j) 20-; 25-; 30-; ...; 55-. (Measurements are given to the nearest
   (k) Above 20 but not exceeding 25;
                                        30;
                30 "
                                        35; etc.
   (1) 20 and above but below 25;
       25 " "
        30 "
                                   35: etc.
  (m) 20-24; 25-29; 30-34; ... (Age last birthday)
Solution (a), (f), (h), (k), (l), (m):-
                20-25; 25-30; 30-35; ...
   (b), (g), (i):
                19.5-24.5; 24.5-29.5; 29.5-34.5; ...
    (c) 19.95-24.95; 24.95-29.95; 29.95-34.95; ...
    (d) 19.995-24.995; 24.995-29.995; 29.995-34.995; ...
    (e) 20.005-25.005; 25.005-30.005; 30.005-35.005; ...
    (j) 19\frac{15}{16} - 24\frac{15}{16}; 24\frac{15}{16} - 29\frac{15}{16}; 29\frac{15}{16} - 34\frac{15}{16}; ...
```





CONSTRUCTION OF FREQUENCY DISTRIBUTION

Example 4.4 Describe the various steps in the construction of a frequency distribution from unclassified data. [C.U., M.Com. '69, '74; B.A.(Econ) '72, '74]

Solution The steps in the construction of a frequency distribution from ungrouped data are as follows:-

- Find the largest and the smallest observations in the given data, and then calculate the range, i.e. the difference between them.
- 2. Divide the range into a suitable number of class intervals, by means of class limits. The number of class intervals should not ordinarily be less than 5 nor more than 15, depending on the number of observations available. The more the number of observations, the larger will be the number of class intervals. Class limits should be so chosen that most of the observations lie near the class marks. The class intervals should preferably be of the same width. In special circumstances, class intervals of unequal width may also be used.
- The number of observations failling in each class interval, i.e. class frequency, is determined by tally marks (see Table 4.19).
- A table is now prepared showing the class intervals in the first column and the corresponding class frequencies in the second column. This is the required frequency distribution.



Example 4.5 Discuss the problems in the construction of a frequency distribution from raw data, with particular reference to the choice of number of classes and class limits. [I.C.W.A., Jan '72; C.A., May '72; C.U., M.Com. '71; B.A. (Econ)'64]

Solution In constructing a grouped frequency distribution, the two main problems involved are: (1) choice of the number of classes, and (b) choice of the class limits. Although there are no hard and fast rules in this regard, it is generally agreed that the number of classes should neither be very large (because, in that case the frequency distribution will be very lengthy), nor very small (because, in that case the true pattern of the distribution of observations will be obscured). As a working rule, this number should lie between 5 and 15, depending on the number of observations available. When the number of observations is small, some authors suggest the use of Sturges' formula

$$n = 1 + 3.3 (\log N)$$

where n denotes the number of classes and N the total frequency. In practice, the number of classes is fixed from other considerations depending on the particular situation.

The second problem, viz. the choice of class limits is to some extent dependent on the number of classes. For this purpose, we first find the maximum and the minimum of the observations and calculate the 'range', i.e. the difference between these two. If we like to have classes of equal width, then the approximate width of the classes may be obtained on dividing the range by the number of such classes. The lower limit of the first class may not coincide with the minimum observation. In most cases we prefer to have classes of width 5 or its multiple, e.g. 15-19, 20-24 etc., or 60.01-80.00, 80.01-100.00, 100.01-120.00 etc., one of the class limits being preferably a multiple of 5. Again, sometimes it is found that some particular values of the variable occur more frequently than others. In such cases, the class limits should be so chosen that the highly frequent values lie near the class marks. This will minimise the errors of grouping (see Section 7.6) in the calculation of various statistical measures like standard deviation, moments, etc. when class intervals are represented by their class marks.



Example 4.6 You are given below the wages paid to some workers in a small factory. Form a frequency distribution with class interval 10 paise:

and the owner of the second second	-	-		Wages in I	Rs.			
1.10	1.13	1.44	1.44	1.27	1.17	1.98	1.36	1.30
1.27	1.24	1.73	1.51	1.12	1.42	1.03	1.58	1.46
1.40	1.21	1.62	1.31	1.55	1.33	1.04	1.48	1.20
1.60	1.70	1.09	1.49	1.86	1.95	1.51	1.82	1.42
1.29	1.54	1.38	1.87	1.41	1.77	1.15	1.57	1.07
1.65	1.36	1.67	1.41	1.55	1.22	1.69	1.67	1.34
1.45	1.39	1.25	1.26	1.75	1.57	1.53	1.37	1.59
1.19	1.52	1.56	1.32	1.81	1.40	1.47	1.38	1.62
1.76	1.28	1.92	1.46	1.46	1.35	1.16	1.42	1.78
1.68	1.47	1.37	1.35	1.47	1.43	1.66	1.56	1.48

[C.A., May '67]

Solution From the given observations, we find that

Maximum value = 1.98 Minimum value = 1.03

The width of the classes is to be 10 paise (Note that in the question 'class interval 10 paise' implies that the width of the classes is to be 10 paise). Let us take the class limits 1.01–1.10; 1.11–1.20; 1.21–1.30; ...; 1.91–2.00. These are shown in a tally sheet (Table 4.19) and the given observations are represented by tally marks one by one against the appropriate classes. The tally marks are shown is groups of five to facilitate counting, every fifth tally mark in a class interval being placed across the preceding four.

Table 4.19 Tally Sheet

Class Limits	Tally Marks	Frequency
1.01-1.10	M	5
1.11-120	MIII	7
1.21-130	Mi Mi	10
1.31-1.40	M M M	15
1.41-1.50	M M M III	18
1.51-160	MI MI IIII	14
1.61-1.70	MIIII	9
1.71-1.80	M	5
1.81-190	S un	4
1.91-2.00	ui	3
Total		90

The required frequency distribution is shown below:

Table 4.20 Frequency Distribution of Wages

Wages (Rs)	Frequency
1.01 – 1.10	5
1.11 – 1.20	7
1.21 – 1.30	10
1.31 – 1.40	15
1.41 – 1.50	18
1.51 – 1.60	14
1.61 – 1.70	9
1.71 - 1.80	5
1.81 – 1.90	4
1.91 – 2.00	3
Total	90



Example 4.7 In a test run of 50 cars of the same model, the average mileage covered per gallon of fuel consumed by each car was as follows:

18.6	33.4	25.3	27.8	30.6	31.9	33.0	26.3	24.9	29.4
20.0	26.2	28.1	33.1	37.5	22.5	39.1	32.9	33.8	52.6
32.5 26.9	34.6 23.8	32.7	9.5	38.5	29.6	25.3	49.5	30.1	27.9
34.1	38.6	36.0 25.9	38.0 40.6	27.5 53.8	32.3 29.3	34.2	23.1	34.7	29.0

Arrange the data in a frequency table with six classes. Clearly indicate the class boundaries and the class marks.

[B.U., B.A.(Econ) '73]

Solution [Note: (i) Since the given data are rounded to one decimal place, the class limits must also be shown in one decimal (see Table 4.21). Class limits in whole numbers like 9-16, 17-24, ..., or in two decimals like 9.01-16.50, 16.51-24.00, ... will not do here.

(ii) Unlike Example 4.6, the width of classes is not specified here. So the common width has to be determined from the data, keeping in view the total number of classes in the proposed frequency distribution.]

In the given data, Maximum value = 53.8 and Minimum value = 9.5, so that Range = 53.8–9.5 = 44.3. If we are to use 6 classes of equal width, the common width will be approximately 44.3 + 6 = 7.38 = 7.5 (suppose). Using the class limits 9.1-16.5, 16.6-24.0, ... 46.6-54.0, and by tally marking, the frequency table has been obtained as shown below:

Table 4.21 Frequency Table showing Average Mileage Covered per Gallon of Fuel by 50 Cars

Class Limits	Frequency	Class Boundaries	Class Mark
9.1-16.5	1	9.05-16.55	12.8
16.6-24.0	5	16.55-24.05	20.3
24.1-31.5	18	24.05-31.55	27.8
31.6-39.0	21	31.55-39.05	35.3
39.1-46.5	2	39.05-46.55	42.8
46.6~54.0	3	46.55-54.05	50.3
Total	50		





CUMULATIVE FREQUENCY DISTRIBUTION

In statistical investigations, sometimes we are interested in the number of observations smaller than (or greater than) a given value. In such cases, or chief concern is the accumulated frequency upto (or above) some value of the variable. This accumulated frequency is known as 'cumulative frequency'.

Cumulative Frequency corresponding to a specified value of the variable may be defined as the number of observations smaller than (or greater than) that value.

The number of observations 'upto' a given value is called *less-than cumulative* frequency; and the number of observations 'greater than' a value is called the more-than cumulative frequency. 'Cumulative frequency' only refers to the less-than type. When a frequency distribution is given, the cumulative totals of frequencies give the cumulative frequencies. When a grouped frequency distribution relates to a variable of the continuous type, the cumulative frequencies calculated therefrom must be shown against the class boundary points (i.e. end-points of classess). Cumulative frequency expressed as a percentage of total frequency, is known as Cumulative Percentage.

A table showing the cumulative frequencies against values of the variable systematically arranged in increasing (or decreasing) order is known as Cumulative Frequency Distribution. If cumulative percentages are shown, instead of cumulative frequencies, the table is called Cumulative Percentage Distribution.

Example 4.8 What do you mean by a cumulative frequency distribution? Point out its special advantages and uses. [C.A., May '67; I.C.W.A., Jan. '71]

Solution Cumulative Frequency Distribution is a statistical table which shows the values of the variable and the corresponding cumulative frequencies. It can be derived from a grouped frequency distribution by writing down the consecutive class boundary points and noting the number of observations less than (or more than) each class boundary point. A cumulative frequency distribution is of advantage in such studies as the number or percentage of observations less than (or more than) a specified value of the variable.

Uses—Cumulative frequency distribution is used (i) to find the number of observations less than (or more than) any given value; (ii) to find the number of observations falling between any two specified values of the variable; (iii) to find median, quartiles, deciles and percentiles (Section 5.18) graphically; i.e. in general, to find the value of the variable below (or above) which a specified number or percentage of the total frequency lies.

Example 4.9 Construct (a) the cumulative frequency distribution, and (b) the grouped frequency distribution, from the following data:

Value	Frequency
Less than 10	4
" " 20	16
" " 30	40
" " 40	76
" " 50	96
" " 60	112
" " 70	120
" " 80	125

Solution The cumulative frequency distribution shows the values of the variable (class boundaries) and the corresponding cumulative frequencies (less-than type):

Table 4.22 Cumulative Frequency Distribution

Value	Cumulative Frequency (less-than)
10	History and the second
20	16
30	40
40	76
50	96
60	112
- 70	120
80	125

The grouped frequency distribution shows the values of the variable in class intervals and the corresponding class frequencies:

Table 4.23 Grouped Frequency Distribution

Value	Frequency		
0 – 10	4		
10-20	12		
20 – 30	24		
30 – 40	36		
40 – 50	20		
50-60	16		
60 – 70	8		
70 – 80	5		
Total	125		





Example 4.10 From the following data, calculate

- (a) "less-than" and "more than" cumulative frequencies,
- (b) cumulative frequency distribution, and
- (c) cumulative percentage distribution:

Age (years)	Frequency	
15 – 19	37	
20-24	81	
25-29	43	
30 – 34	24	
35 – 44	9	
45 – 59	6	
Total	200	

Solution [Working note: The class boundaries are 14.5-19.5, 19.5-24.5, etc. (Table 4.4). The class boundary points are therefore, 14.5, 19.5, 24.5, 29.5, 34.5, 44.5, 59.5. There is no observation below 14.5, and hence its cumulative frequency is 0; the frequency below 19.5 is 37; the frequency below 24.5 is (37 + 81), i.e. 118; the frequency below 29.5 is (37 + 81 + 43), i.e. 161; and so on. The "less-than" cumulative frequencies corresponding to the above class boundary points are therefore 0, 37, (37 + 81), (37 + 81 + 43), (37 + 81 + 43 + 24),

The "more-than", cumulative frequencies are calculated by cumulating the class frequencies from the higher end of the values. Corresponding to the class boundary points in reverse order 59.5, 44.5, 34.5, 29.5, 24.5, 19.5, 14.5 these are respectively 0, 6, (6 + 9), (6 + 9 + 24), (6 + 9 + 24 + 43 + 81), (6 + 9 + 24 + 43 + 81 + 37), i.e. 0, 6, 15, 39, 82, 163, 200.

It should be noted that the "less-than" cumulative frequency corresponding to the highest class boundary point and the "more-than" cumulative frequency corresponding to the lowest class boundary point must be equal to the total frequency N. Also, the sum of the "less-than" and the "more-than" cumulative frequencies corresponding to any value is equal to the total frequency]















Table 4.24 Cumulative Frequencies

	Age (years)	Cumulative Frequency	
		"less than"	"more-than"
en accused as	14.5		200 a N
	19.5	37	163
	24.5	118	82
	29.5	161	39
	34.5	185	15
	44.5	194	6
	39.5	200 ≈ N	0

The cumulative frequency distribution shows the values of the variable (class boundary points) and the corresponding cumulative frequencies (less-than type) side by side: