Frequency distributions

Classification

When raw data are collected, they are distributed into classes, or categories, ot intervals and the number of times the items appear in a class-interval is known as the class frequency.

Procedure for tabulating a frequency distribution

(i) Fix the magnitude of the class interval

(ii) If possible allow for a maximum of 20-25 class-intervals

(iii) Classify the items according to (i) and (ii).

(iv) Draw up a table showing te frequency of each class-interval

Raw data:

7.1, 7.5, 8.2, 9.2, 5.4, 6.1, 2.3, 5.4, 2.8, 5.8, 8.5, 4.2, 6.8, 4.8, 5.8, 1.8,
5.8, 5.9, 7.8, 4.7, 6.8, 9.9, 10.2, 15.8, 12.8, 7.8, 4.5, 6.7, 5.7, 6.6, 8.5,
7.5, 4.1, 5.8, 2.8, 6.8, 5.5, 5.4, 6.7, 8.7, 5.7, 12.4, 7.7, 5.8, 8.2, 5.6

Raw data:

7.1, 7.5, 8.2, 9.2, 5.4, 6.1, 2.3, 5.4, 2.8, 5.8, 8.5, 4.2, 6.8, 4.8, 5.8, 1.8,
5.8, 5.9, 7.8, 4.7, 6.8, 9.9, 10.2, 15.8, 12.8, 7.8, 4.5, 6.7, 5.7, 6.6, 8.5,
7.5, 4.1, 5.8, 2.8, 6.8, 5.5, 5.4, 6.7, 8.7, 5.7, 12.4, 7.7, 5.8, 8.2, 5.6

Minimum item = 1.8

Maximum item = 15.8

Number of class-intervals = 15

Class-interval length = 1

C-I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5	14.5	15.5
f	1	3	0	5	13	7	6	5	2	1	0	2	0	0	1
cf	1	4	4	9	22	29	35	40	42	43	43	45	45	45	46
rf	0,022	0,065	0	0,109	0,283	0,152	0,13	0,109	0,043	0,022	0	0,043	0	0	0,022
rcf	0,022	0,087	0,087	0,196	0,478	0,63	0,761	0,87	0,913	0,935	0,935	0,978	0,978	0,978	1

C-I class interval

R = Representative
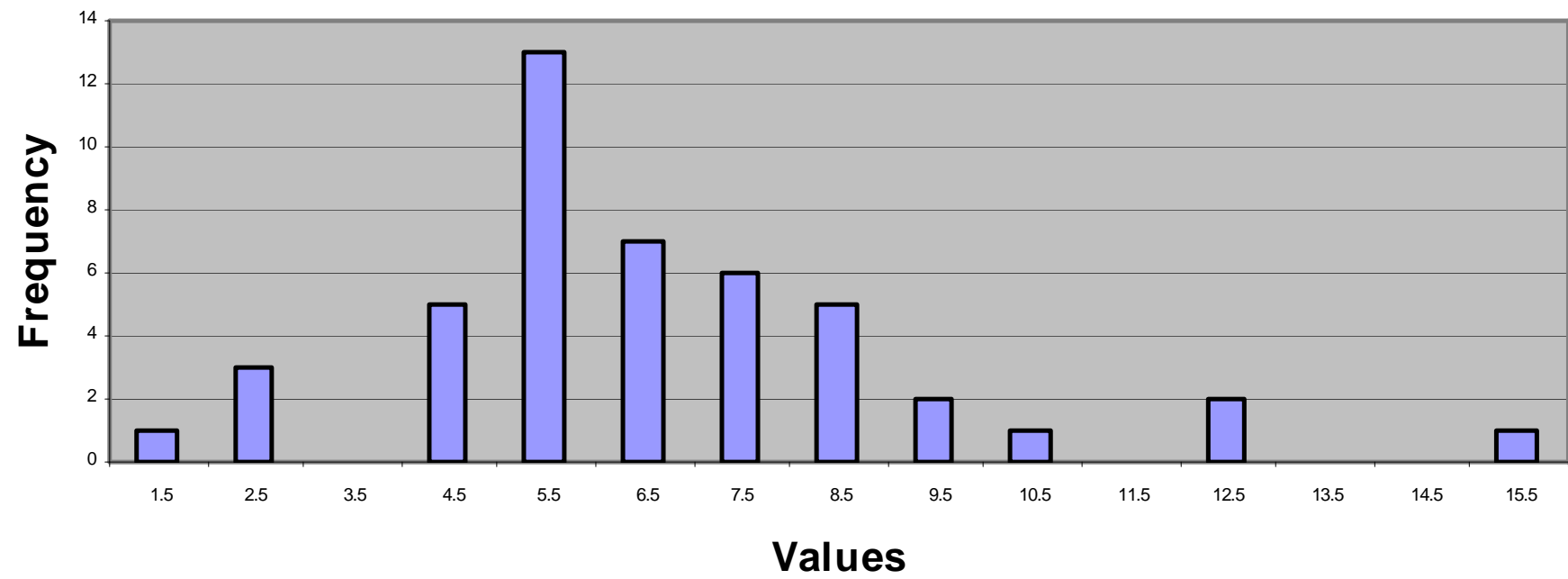
f = Frequency

cf = Cummulative frequency

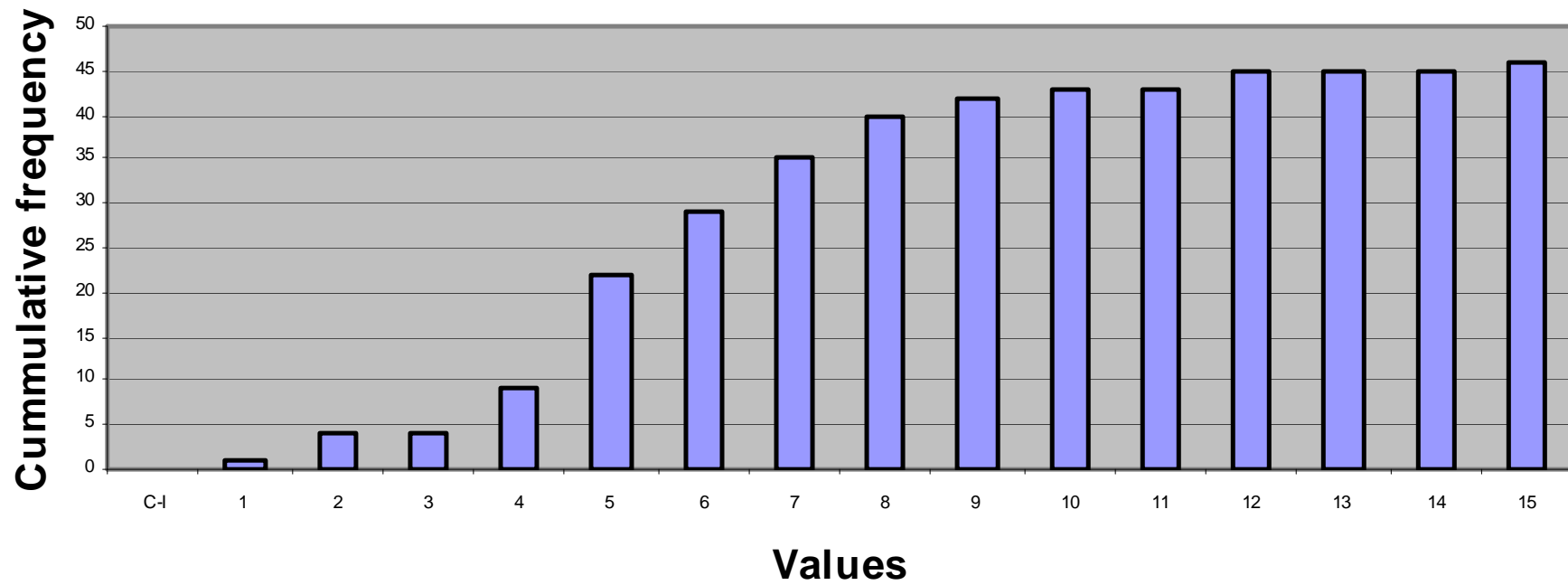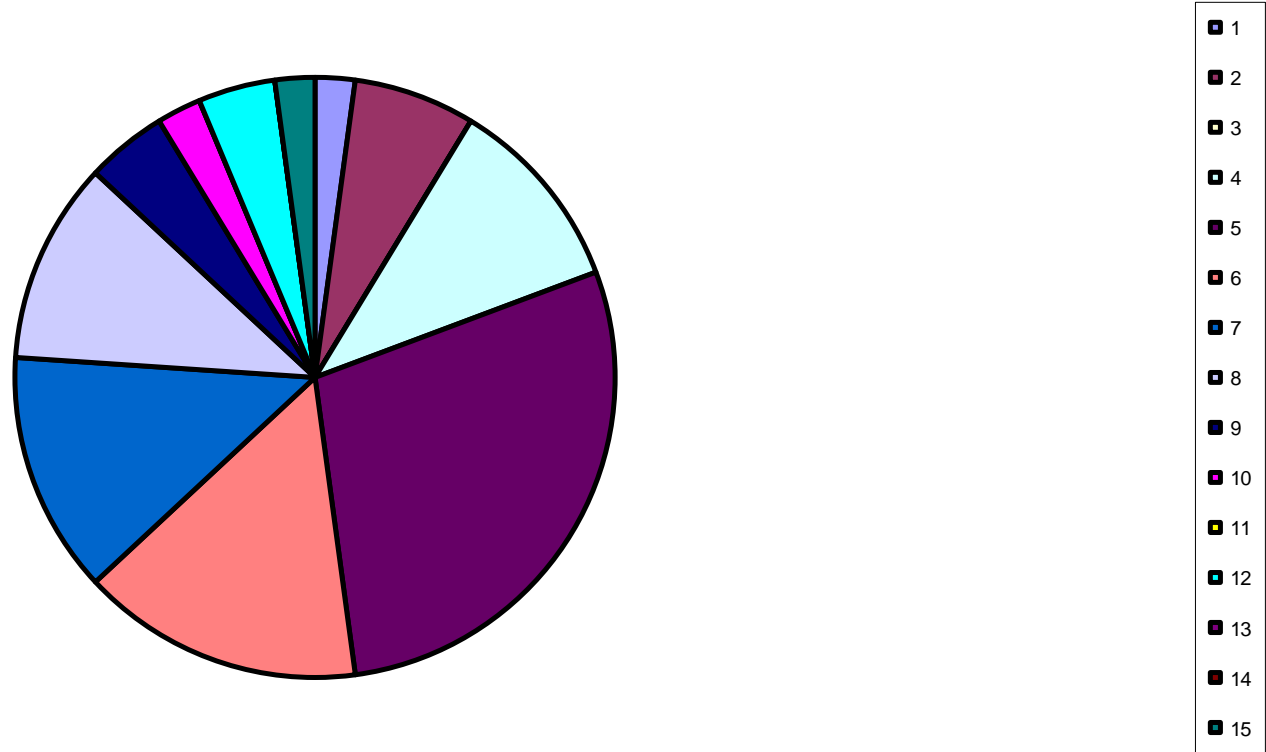rf = Relative frequency

rcf = Relative cummulative frequency

Frequency distribution histogram

Cummulative frequency distribution histogram

Relative frequency distribution pie-chart

Legend:
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15

With a given set of raw data $x_1, x_2, \ldots, x_n$

or a frequency distribution table

r	r_1	r_2	...	r_k
f	f_1	f_2	...	f_k

We can calculate various data characteristics such as

range, arithmetic mean, median, mode, quantiles, variance, standard deviation

Range

Range is the difference between the magnitude of the largest and the smallest of data items.

The arithmetic mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{\sum_{j=1}^{k} x_j f_j}{\sum_{j=1}^{k} f_j}$$

Median

Arrange the items of a distribution in order of magnitude strating with either the largest or the smallest, then

(i) if the number of items is odd, the median is the value of the middle item

(ii) if the number of items is even, the median is the arithmetic mean of the two middle items

The mode

The mode of a distribution is the most frequent, or most 'popular' item. It may not exist and, even if it does, it may not be unique

Quantiles

median - divides a distribution into two equal parts

quartiles Q_1, Q_2, Q_3, Q_4 - divide a distribution into four equal parts

deciles D_1, ..., D_{10} - divide a distribution into ten equal parts

percentiles P_1, ..., P_{100} - divide a distribution into a hundred equal parts

Q_2 is the median

Variance, standard deviation

$$s^2 = \frac{\sum\limits_{i}^{n}(x_i - \bar{x})^2}{n}$$

$$s'^2 = \frac{\sum\limits_{i}^{n}(x_i - \bar{x})^2}{n-1}$$

$$s = \sqrt{\frac{\sum\limits_{i}^{n}(x_i - \bar{x})^2}{n}}$$

$$s' = \sqrt{\frac{\sum\limits_{i}^{n}(x_i - \bar{x})^2}{n-1}}$$