Chapter 7

CHANCE AND DATA

Statistics involves the collection and organisation of information (data) so that:

(a) large masses of information can be easily analysed, and

(b) predictions can be made, based on the collected information.

Tables and graphs allow information to be presented in a clear, concise form. The information can also be readily analysed from the table or graph.

7.1 Simple distributions

7.1.1 Some definitions and formulae

Range = highest score – lowest score

Mode = most common score or score with the highest frequency

Median = middle score when scores are arranged in *ascending* order

$$\text{Mean} = \text{average} = \frac{\text{total of the scores}}{\text{number of scores}}$$

$$= \frac{\sum xf}{\sum f}$$

7.1.2 A note on notation

x	score
f	frequency
c.f.	cumulative frequency
xf	(score) × (frequency)
$\sum$	sum
$\sum f$	sum of frequency column
$\sum xf$	sum of xf column
$\bar{x}$	mean

Examples

(a) The results of 25 Year One students in a weekly spelling test are:

7	8	9	10	4	7	6
4	5	3	8	5	6	7
9	9	8	7	6	5	7
8	6	4	9			

(i) Construct a frequency distribution table to enable you to answer the following questions:

(ii) How many students scored:
- (A) 8 marks.
- (B) 8 or less marks.
- (C) less than 8 marks.
- (D) more than 8 marks.

(iii) What percentage of students scored:
- (A) 8 marks.
- (B) 8 or less marks.
- (C) less than 8 marks.
- (D) more than 8 marks.

(iv) Find the range and the mode.

(v) Calculate the mean.

(vi) Find the median score.

(vii) Construct:
- (A) a frequency histogram.
- (B) a cumulative frequency histogram.

SOLUTION

(i) Frequency distribution table

Score (x)	Tally	Frequency (f)	Cumulative frequency (c.f.)	Score × frequency xf
3	|	1	1	3 ⇐ (3 × 1)
4	|||	3	4	12 ⇐ (4 × 3)
5	|||	3	7	15 ⇐ (5 × 3)
6	|||\	4	11	24
7	|||\ |	5	16	35
8	|||\	4	20	32
9	|||\	4	24	36
10	|	1	25	10
	Σf	25	Σxf	167

(ii) (A) 4 (from frequency column)
(B) 20 (from c.f. column)
(C) 18 [same as 7 or less]
(D) 25 − 20 = 5 [total less (8 or less)]

(iii) (A) Percentage $= \frac{4}{25} \times 100\%$
$= 16\%$

(B) Percentage $= \frac{20}{25} \times 100\%$
$= 80\%$

(C) Percentage $= \frac{16}{25} \times 100\% = 64\%$

(D) Percentage $= \frac{5}{25} \times 100\% = 20\%$

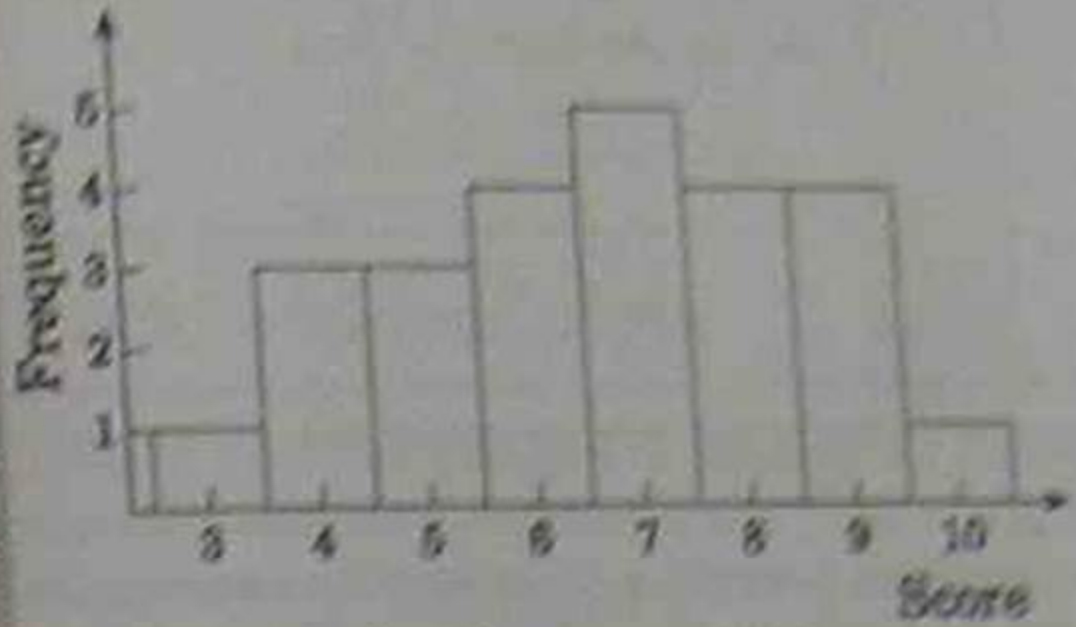
(iv) Range
= highest score − lowest score
$= 10 - 3 = 7$
Mode
= score with highest frequency
= 7 (frequency is 5)

(v) Mean $= \frac{\sum xf}{\sum f}$
$= \frac{167}{25}$
$= 6.68$

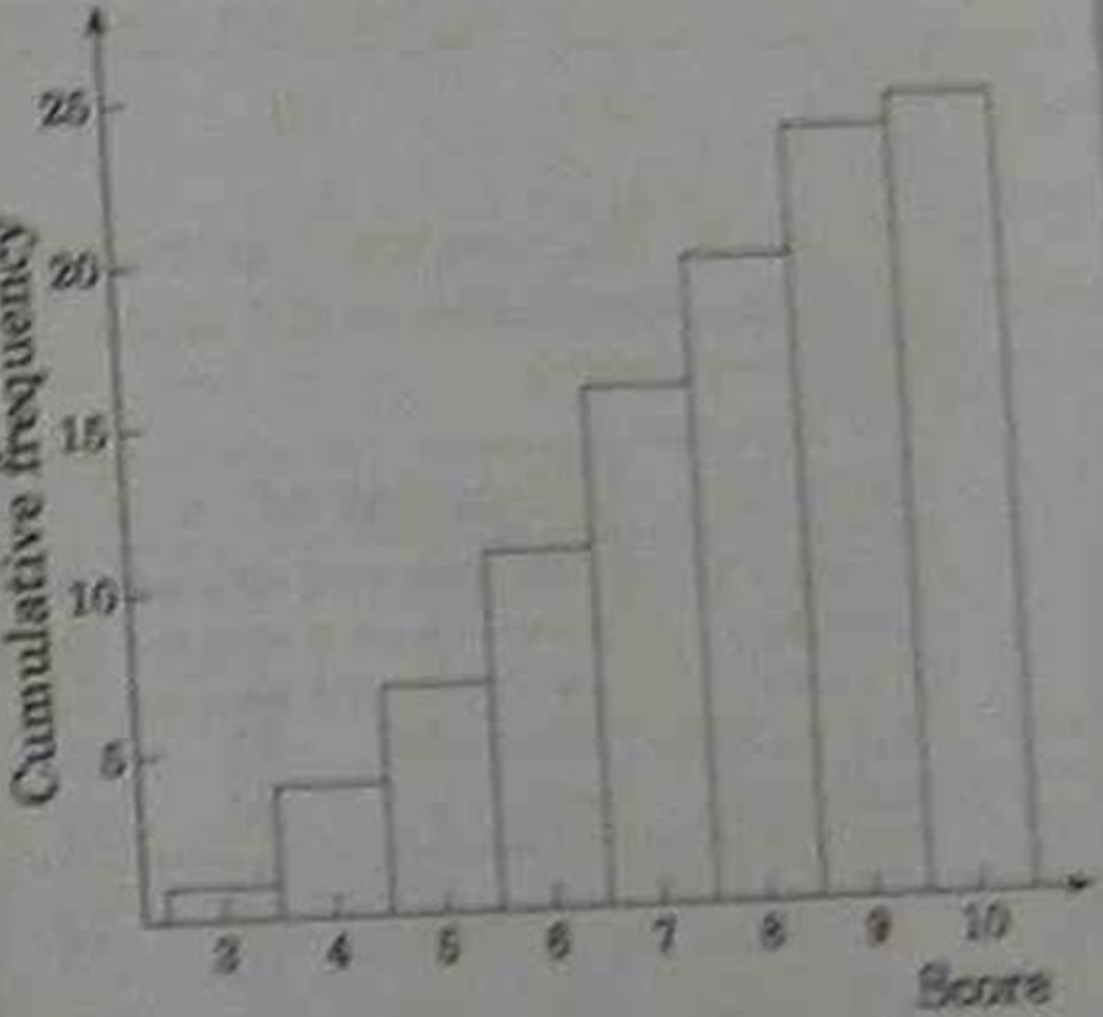
(vi) To find the median, we need the middle score. This is the 13th score. From the c.f. column, the 13th score is a 7. (The 12th, 13th, 14th, 15th, 16th scores are all 7.)

> When we have 25 scores the 13th is the middle score, that is, 12 before it and 12 after it.

(vii) (A) Histogram (Column graph)

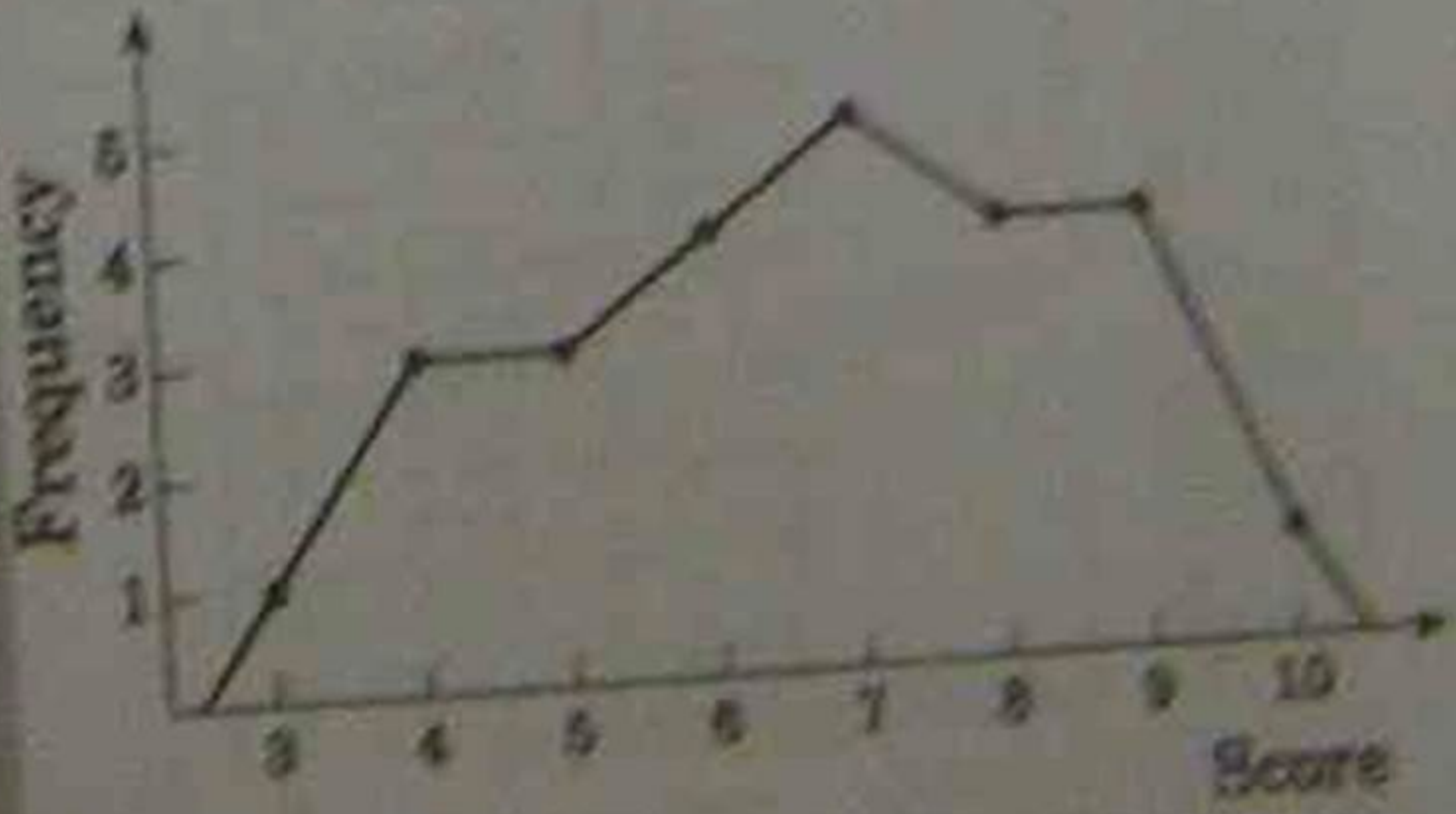


(B) Cumulative frequency histogram



Note that if a frequency polygon had been required, the graph would be as follows:

Frequency polygon



This can also be achieved by joining the midpoints of the tops of the columns in the histogram in (A).

(b) Sometimes, when there are only a few scores, a table is not necessary.

Remember your calculator will calculate the mean. For example: Gregory, over his first six innings, managed scores of 18, 32, 12, 12, 25 and 12.

(i) Find the range of these scores.

(ii) Write down the modal score.

(iii) Calculate the mean and the median.

(iv) If in his next three innings he scores 17, 15 and 25, find the new range, mode, mean and median.

(v) How many runs must he score in his tenth innings to have an overall average of 18?

SOLUTION

(i) Range

= highest score − lowest score

= 32 − 12

= 20

(ii) Mode = most common score

= 12

(iii) Mean = $\frac{\text{total of scores}}{\text{number of scores}}$

$= \frac{13 + 32 + 12 + 12 + 20) + 12}{6}$

$= \frac{106}{6}$

= 17.7 (one decimal place)

Median = middle score when scores are in ascending order

The scores in ascending order are:

12, 12, 12, 18, 20, 32

Middle scores

$$\text{Median} = \frac{12 + 18}{2}$$

$$= 15$$

When there is an even number of scores, the median will be the average of the two middle scores, that is, half-way between the two middle scores.

(iv) The scores are now 18, 32, 12, 12, 20, 12, 17, 15 and 20.

Range $= 32 - 12 = 20$ [unchanged]
Mode $= 12$ [unchanged]
Mean

$$= \frac{18+32+12+12+20+12+17+15+20}{9}$$

$$= \frac{158}{9} = 17.6 \text{ (one decimal place)}$$

The scores in ascending order are:
12, 12, 12, 15, 17, 18, 20, 20 and 32.

↑

Middle score

When there is an odd number of scores there is a single middle score.

The mean has dropped marginally while the median has risen.

(v) For Gregory to average 18 over ten innings, his total for ten innings would be (10×18) runs, that is, 180 runs. Over nine innings, he has scored 158 runs, thus he would have to score (180 – 158) runs, that is, 22 runs in his tenth innings

7.1.3 Dot plots

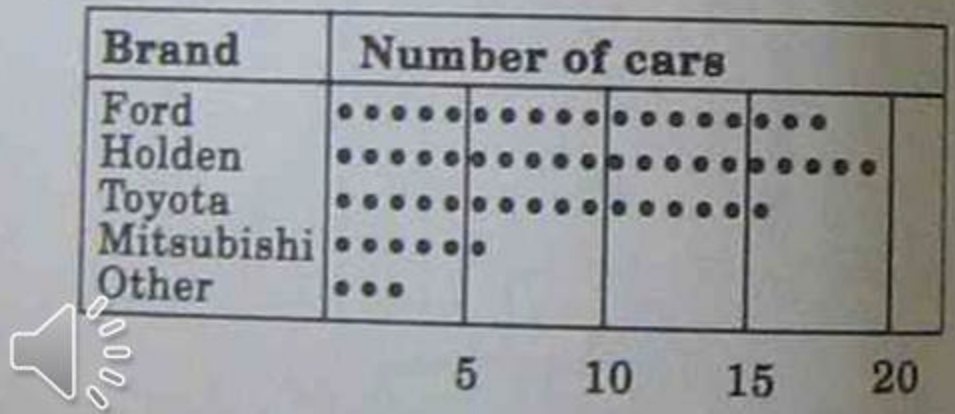
A dot plot is an alternative method of displaying information. It is simpler to draw and simpler to read than a histogram.

Example:

The type of cars passing along Macquarie Street over a half an hour period was noted and the results recorded in the following table:

Brand	Frequency
Ford	18
Holden	20
Toyota	16
Mitsubishi	6
Other	3
Total	63

A dot diagram was constructed from the information in the table.



A dot diagram can also be used in the same way as a tally to organise raw data.

A dot diagram can also be used in the same way as a tally to organise raw data.

Example:

Consider the results of 25 Year One students in a weekly spelling tests:

7	8	9	10	4
7	6	4	5	3
8	5	6	7	9
9	8	7	6	5
7	8	6	4	9

Firstly, a rough dot plot is done

The numbers are transferred to the diagram as dots in the order that they occur.

This is then transferred to a frequency distribution table

Score	Frequency
3	1
4	3
5	3
6	4
7	5
8	4
9	4
10	1
	25

The rough dot plot serves as a tally as well as giving an indication of the size of the distribution and other features such as outliers.

7.1.4 Stem and leaf plots

This is another method of displaying numerical data. The leaf is the FINAL digit of a number where the stem is the FIRST digit or digits. The leaf is always a single digit. The stem can contain any number of digits.

For example consider these numbers: 68, 647, 24317. If these should be organised into stem and leaf then:

68	6 stem	8 leaf
647	64 stem	7 leaf
24 317	2431 stem	7 leaf

Example:

The results in a mathematics test out of 50 were:

19	48	36	40	31	22
18	27	18	20	18	36
49	50	45	13	9	17
22	31	39	26	28	30
44	8	23	19	46	33

Draw a stem and leaf plot for this data.

Initial plot

Stem is tens digit

Leaf is units digit

Stem	Leaf
0	9 8
1	9 8 8 8 3 7 9
2	2 7 0 2 6 8 3
3	6 1 6 1 9 0 3
4	8 0 9 5 4 6
5	0

this row represents the scores 48, 40, 49, 45, 44, 46

Note that in the initial plot the numbers are entered as they occur.

Refined plot

Stem	Leaf
0	8 9
1	3 7 8 8 8 9 9
2	0 2 2 3 6 7 8
3	0 1 1 3 6 6 9
4	0 4 5 6 8 9
5	0

The digits in the leaf are arranged in increasing order in order to enable greater understanding of the scores.

A stem and leaf plot allows the smallest and largest numbers to be seen, as well as the size of the distribution and any clustering of data. It gives an indication of the groupings to be used in a grouped table and makes the tallying easier.

An application of the stem and leaf plot is to find range, mode, ... plot.

Example

Consider the stem and leaf plot drawn to illustrate the results of 30 students in a ... ematics test. Find the range, mode, mean and median.

SOLUTION

Stem	Leaf	Number of leaves f	$f \times$ stem	Sum of leaves
0	89	2	0	17
1	3788899	7	70	52
2	0223678	7	140	28
3	0113669	7	210	26
4	045689	6	240	32
5	0	1	50	0
	Σ	30	710	155

- Range = Highest score – Lowest score = 50 – 8 = 42
- Mode = score(s) with highest frequency = 18

Look for digit occuring the most against each stem

- $$\text{Mean} = \frac{\text{Sum of } (\lceil f \times \text{stem} \rceil \text{ and } \lceil \text{sum of leaves} \rceil)}{\text{Total } f}$$

$$= \frac{710 + 155}{30}$$

$$= \frac{865}{30}$$

$$\approx 28.8 \quad \text{(to 1 decimal place)}$$

$$\bar{x} = \frac{\sum (f \times \text{stem}) + \sum \text{Sum of leaves}}{\sum f}$$

- Median is the middle score — cross numbers off in the leaf column one from start and one from end until only a single number (odd number) or two numbers (even number). Median is the single number or average of the two remaining numbers.

and then to here ⇓

Stem	Leaves
0	~~8~~ ~~9~~
1	~~3~~ ~~7~~ ~~8~~ ~~8~~ ~~8~~ ~~9~~ ~~9~~
2	~~0~~ ~~2~~ ~~2~~ ~~3~~ ~~6~~ ~~7~~ 8
3	~~0~~ ~~1~~ ~~1~~ ~~3~~ ~~6~~ ~~6~~ ~~9~~
4	~~0~~ ~~4~~ ~~5~~ ~~6~~ ~~8~~ ~~9~~
5	~~0~~

⇑ Start here

27 and 28 are left
∴ the median is

$$\frac{27+28}{2} = 27.5$$

Median = 27.5

Range = 42 Mode = 18 Mean = 28.8 Merdian = 27.5