PART C

Chapter 5

Descriptive statistics
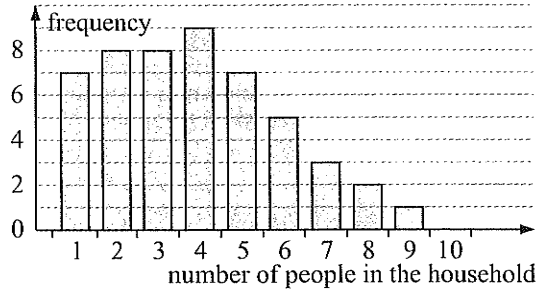
4 A randomly selected sample of households has been asked, 'How many people live in your household?' A column graph has been constructed for the results.

    a How many households gave data in the survey?

    b How many of the households had only one or two occupants?

    c What percentage of the households had five or more occupants?

    d Describe the distribution of the data.

5 The number of matches in a box is stated as 50 but the actual number of matches has been found to vary. To investigate this, the number of matches in a box has been counted for a sample of 60 boxes:

51 50 50 51 52 49 50 48 51 50 47 50 52 48 50 49 51 50 50 52
52 51 50 50 52 50 53 48 50 51 50 50 49 48 51 49 52 50 49 50
50 52 50 51 49 52 52 50 49 50 49 51 50 50 51 50 53 48 49 49

    a What is the variable in this investigation?

    b Is the data continuous or discrete numerical data?

    c Construct a frequency table for this data.

    d Display the data using a bar chart.

    e What percentage of the boxes contained exactly 50 matches?

6 Revisiting **The Pea Problem**.   For the *With fertiliser* data:

    a Organise the data in a tally-frequency table.

    b Draw a column graph of the data.

    c Are there any outliers?

    d What evidence is there that the fertiliser 'increases the number of peas in a pod'?

    e Can it be said that the fertiliser will increase the farmer's pea crop and his profits?

C   GROUPED DISCRETE DATA

A local kindergarten is concerned about the number of vehicles passing by between 8.45 am and 9.00 am. Over 30 consecutive week days they recorded data.

The results were: 27, 30, 17, 13, 46, 23, 40, 28, 38, 24, 23, 22, 18, 29, 16, 35, 24, 18, 24, 44, 32, 52, 31, 39, 32, 9, 11, 38, 24, 32

In situations like this we group the data into **class intervals**.

It seems sensible to use class intervals of length 10 in this case.

The tally frequency table is:

Number of cars	Tally	Frequency
0 to 9	│	1
10 to 19	ﬀ	5
20 to 29	ﬀ ﬀ	10
30 to 39	ﬀ ﬀ	9
40 to 49		4
50 to 59		1
	Total	30

STEM-AND-LEAF PLOTS

A **stem-and-leaf plot** (often called a **stemplot**) is a way of writing down the data in groups. It is used for small data sets.

A stemplot shows actual data values. It also shows a comparison of frequencies.

For numbers with two digits, the first digit forms part of the **stem** and the second digit forms a **leaf**.

For example, for the data value 18,   1 is recorded on the stem, 8 is a leaf value.

For the data value 123, 12 is recorded on the stem and 3 is the leaf.

The **stem-and-leaf plot** is:

Stem	Leaf
0	9
1	73868
2	7384329444
3	085219282
4	6041
5	2

Note: 1 . 7   means 17

The **ordered stem-and-leaf plot** is:

Stem	Leaf
0	9
1	36788
2	2334444789
3	012225889
4	0146
5	2

The ordered stemplot arranges all data from smallest to largest.

Notice that:

- all the actual data is shown
- the minimum (smallest) data value is 9
- the maximum (largest) data value is 52
- the 'twenties' interval (20 to 29) occurred most often.

BACK TO BACK STEM-AND-LEAF PLOTS

This back to back stemplot represents the times for the 100 metre freestyle recorded by members of a swimming squad.

Note: The fastest time for a girl is 33.4 seconds and the slowest time for a girl is 41.1 seconds.
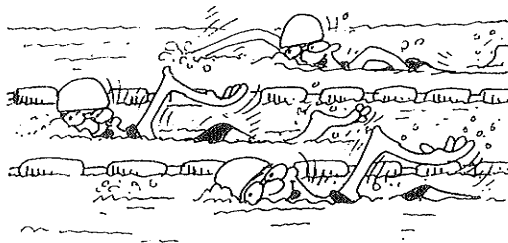
| Girls	Stem	Boys
	32	1
4	33	0227
763	34	13448
87430	35	024799
8833	36	788
7666	37	0
6	38	
0	39	
	40	
1	41	

leaf unit: 0.1 seconds.

COLUMN GRAPHS

A vertical column graph can be used to display grouped discrete data.

For example, consider the local kindergarten data of page **117**.

The **frequency table** is:

Number of cars	Frequency
0 to 9	1
10 to 19	5
20 to 29	10
30 to 39	9
40 to 49	4
50 to 59	1

The **column graph** for this data is:
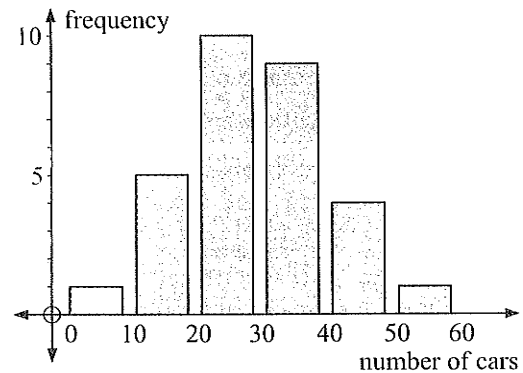
INVESTIGATION 2   TAXI SIR?

Two taxi drivers, Peter and Ivan, are friendly rivals. Each claims that he is the more successful driver. They agree to randomly select 25 days on which they work and record the daily fare totals. The data collected to the nearest dollar was:

Peter   194   99   188   208   95   168   205   196   233
116   132   153   205   191   182   118   140   270
183   155   93   154   190   223   147

Ivan   260   152   127   163   180   161   110   153   139
139   142   161   97   116   129   215   241   160
110   159   147   174   162   158   223

What to do:

1 Produce back-to-back stem plots of this data.

2 Explain why "the amount of money collected per hour" would have been a better variable to use rather than "the daily fare totals".

3 Produce back-to-back stem plots for "the amount of money collected per hour" given below.

Peter ($ per hour)   17.27   11.31   15.72   18.92   9.55   12.98   19.12   16.69   11.68
15.84   12.81   24.03   15.03   12.95   20.09   18.64   18.94   13.92
11.69   15.52   15.21   18.26   12.25   18.59   22.79

Ivan ($ per hour)   23.70   13.30   12.18   14.20   15.74   14.01   10.05   10.05   12.20
13.50   18.64   13.29   12.65   13.54   8.83   11.09   12.29   18.94
20.08   13.84   14.57   13.34   13.44   13.63   14.18

4 Present a brief written report (of about 250 words) to summarise your findings and present your conclusion.

EXERCISE 5C

1 The data set below is the test scores (out of 100) for a Maths test for 50 students.

56 29 78 67 68 69 80 89 92 58 66 56 88 81 70 73 63 74
67 64 62 55 56 75 90 92 47 59 64 89 62 51 87 89 76 59
72 80 95 68 80 64 53 43 61 71 38 44 88 39

   a Construct a tally and frequency table for this data using class intervals 0 - 9,
      10 - 19, 20 - 29, ..., 90 - 100.
   b What percentage of the students scored 80 or more for the test?
   c What percentage of students scored less than 50 for the test?
   d Copy and complete the following:
      More students had a test score in the interval .... than in any other interval.
   e Draw a column graph of the data.

2 a Draw a stem-and-leaf plot using stems 2, 3, 4, and 5 for the following data:
      27, 34, 25, 36, 57, 34, 42, 51, 50, 48, 29, 27, 33, 30, 46, 40, 35, 24, 21, 58
   b Redraw the stem-and-leaf plot from a so that it is ordered.

3 For the ordered stem-and-leaf plot given find:

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

   a the minimum value
   b the maximum value
   c the number of data with a value greater than 25
   d the number of data with a value of at least 40
   e the percentage of the data with a value less than 15.
   f How would you describe the distribution of the data? **Hint:** Turn your stemplot on its side.

4 The test score, out of 50 marks, is recorded for a group of 45 students.

25 28 35 42 44 28 24 49 29 33 33 34 38 28 26
32 34 39 41 46 35 35 43 45 50 30 22 20 35 48
36 25 20 18 9 40 32 33 28 33 34 34 36 25 42

   a Construct a stem-and-leaf plot for this data using 0, 1, 2, 3, 4, and 5 as the stems.
   b Redraw the stem-and-leaf plot so that it is ordered.
   c What advantage does a stem-and-leaf plot have over a frequency table?
   d What is the i highest ii lowest mark scored for the test?
   e If an 'A' was awarded to students who scored 40 or more for the test, what percentage of students scored an 'A'?
   f What percentage of students scored less than half marks for the test?

5 The students of class 11C and class 11P were asked to record the number of hours they spent watching television during a one week period. The results were as follows:

Class 11C							Class 11P					
21	10	18	27	32	13	5	22	23	29	25	18	30
12	14	26	21	20	37	14	15	14	19	29	7	16
9	15	11	27	17	30	12	24	11	12	14	15	6

   Draw a back to back stemplot to represent this data.