Chapter Five.

The statistical investigation process.

Suppose we were asked to investigate

if a particular species of animal is in decline,
if cutting a lawn makes it grow faster,
if sugary drinks cause hyperactiveness,
if social media networks are addictive
if man made pollution is causing global warming,
if females have a quicker reaction time that males,
if female year 8s are better at mental mathematics than male year 8s,
if year 11s are fitter than year 8s,
etc etc

We could implement the **statistical investigation process**, a process that can be described by the following steps:

① Clarify the problem and formulate one or more questions that can be answered with data.
② Design and implement a plan to collect and obtain appropriate data.
③ Select and apply appropriate graphical or numerical techniques to analyse the data.
④ Interpret the results of this analysis and relate the interpretation to the original question and communicate findings in a systematic and concise manner.

In this chapter you will be invited to carry out a statistical investigation that will involve collecting and comparing data across two or more groups to investigate a question.

For example. Consider the following question:

Are year elevens fitter than year eights?.

To investigate this question we would need to collect data about a group of year elevens and about a group of year eights and compare the findings.

For example. Consider the following question:

Do females have quicker reaction times than males?.

To investigate this question we would need to collect data about a group of males and about a group of females and compare the findings.

Both examples involved collecting data across two groups.

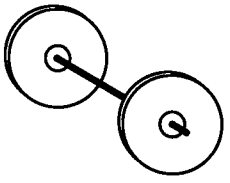
Implementing the statistical investigation process.

Collect data to investigate one of the questions posed below (or negotiate with your teacher to collect data that will involve collecting and comparing data across two or more groups to investigate an alternative question of your own).

You need to think about
- what data you will collect,
- how you will collect it,
- how you will record it.

Present your findings as a report and include in your report:

- ☞ the question you decided to investigate,
- ☞ what data you collected, how you collected it and how you recorded it,
- ☞ the collected data, tabulated and/or graphed as appropriate,
- ☞ your analysis of that data,
- ☞ your conclusions.



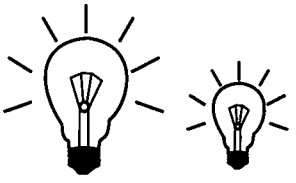
⇨ Are year 11s the fittest year group in the school?

⇨ Do females have quicker reaction times to males?

15 + 17
13 × 5
23 + ? = 71
7 × 5 + 8
7 × (5 + 8)

⇨ Are female year 8s better at mental arithmetic than male year 8s?

⇨ How does your *Mathematics Applications* class compare with other *Mathematics Applications* classes in recent tests



⇨ Do males have bigger heads than females?

⇨ Are left handers generally taller than right handers

Miscellaneous Exercise Five.

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the preliminary work section at the beginning of the book.

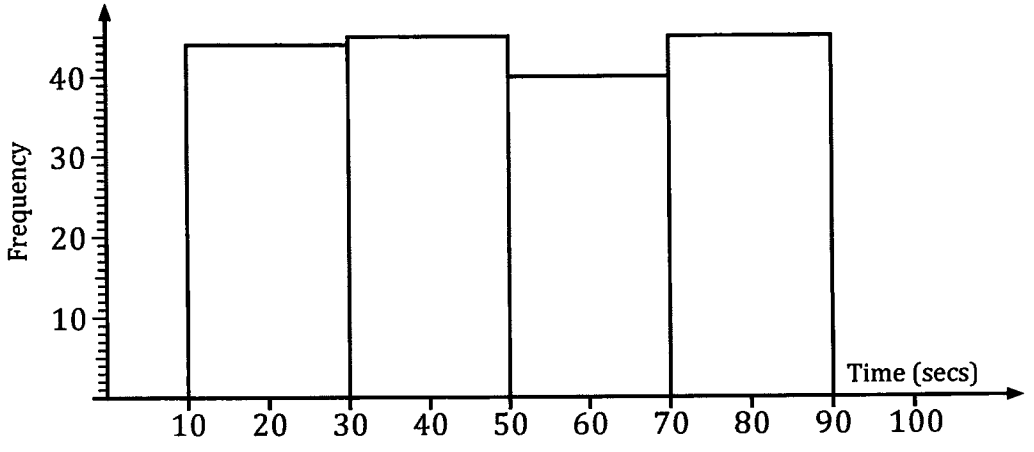
1. The times taken to perform a particular task were recorded for 174 people and the results, in seconds, are shown below:

20·41	30·13	60·34	19·72	81·31	27·14	54·14	70·08	68·73	17·24
51·34	20·32	26·83	84·72	25·12	41·13	24·23	53·24	31·62	67·81
26·14	57·24	29·81	70·02	76·24	55·82	58·21	24·55	24·62	44·14
60·82	26·08	63·54	20·80	10·73	40·21	27·23	64·18	75·64	53·80
29·24	48·24	85·34	57·89	38·13	24·32	39·23	75·00	35·88	23·14
74·18	80·79	11·23	33·15	77·34	68·81	28·14	24·21	16·81	31·54
79·31	38·72	88·23	70·73	69·14	78·85	69·86	70·13	76·88	75·21
18·74	70·54	33·65	38·94	37·24	29·53	48·87	30·82	23·22	45·61
76·31	38·24	56·18	66·84	60·98	62·41	15·71	49·00	64·92	31·70
38·16	48·94	47·32	18·31	39·41	28·32	77·41	39·91	34·73	67·34
74·23	60·93	85·23	74·89	34·13	75·32	61·24	75·87	74·98	23·91
33·24	29·11	36·40	65·11	20·81	51·24	13·75	43·24	29·18	35·24
47·41	71·34	89·10	71·23	76·71	28·71	71·92	67·14	58·90	53·71
79·58	42·09	61·82	85·67	65·01	75·93	61·34	58·32	31·23	47·24
21·32	77·23	88·24	14·63	76·91	42·90	34·89	23·62	53·28	72·81
57·31	18·13	65·23	76·52	48·21	73·24	78·51	23·51	42·64	15·23
42·71	79·22	42·83	36·02	62·13	22·04	30·24	51·89	31·24	64·99
77·14	29·72	21·41	81·25						

The histogram for these results, using the intervals

$10 \le \text{time} < 30$ $30 \le \text{time} < 50$ $50 \le \text{time} < 70$ $70 \le \text{time} < 90$

is shown below.



This histogram suggests that the times are evenly, or *uniformly*, distributed from a low of about 10 seconds to a high of about 90 seconds.

However is this "uniform distribution" really the case?

Draw the histogram for this data using the intervals

$10 \le \text{time} < 20$ $20 \le \text{time} < 30$ $30 \le \text{time} < 40$ etc.,

and comment on your findings.

2. Expand and simplify each of the following:

(a) $2(x+4)+5(x+3)$ (b) $5(2x+3)+2(3x+1)$

(c) $12(x+1)-5(x+2)$ (d) $3(x-4)-2(x-1)$

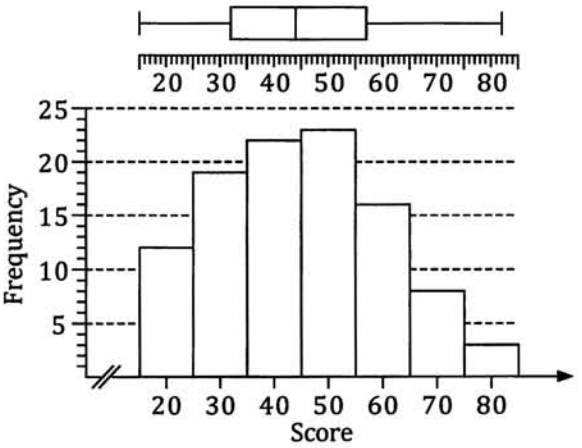
(e) $3(2x+5)-1(x-3)$ (f) $2(3x+2)-(5x+2)$

3. The box plot and the histogram for a set of data are shown on the right.
For this set of data determine

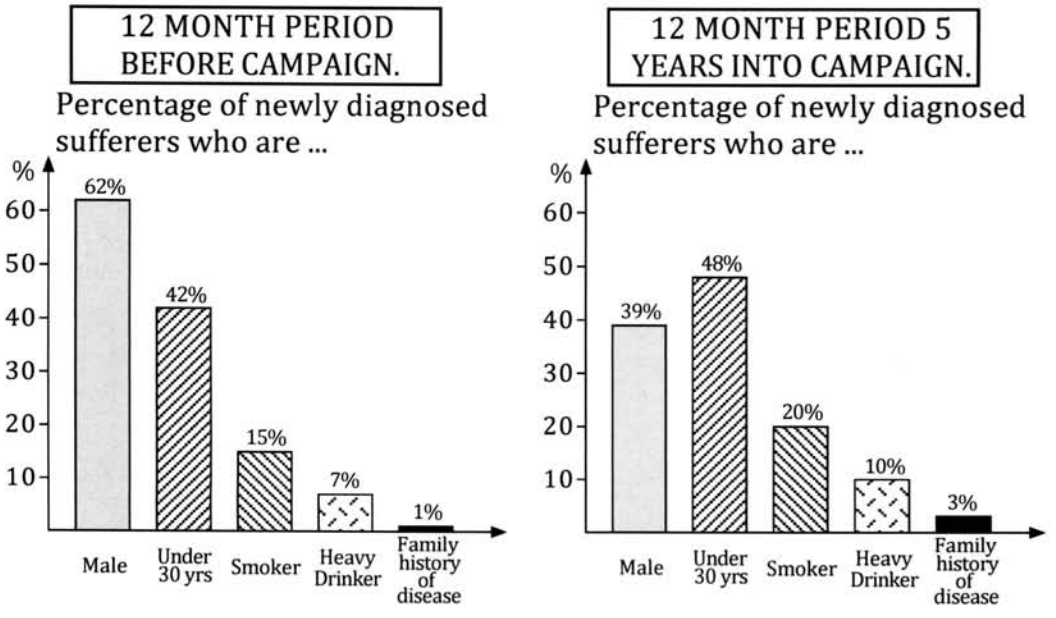
(a) the lowest score,
(b) the median,
(c) the highest score,
(d) the interquartile range,

and use your calculator to determine

(e) an estimate for the mean,
(f) an estimate for the standard deviation.



4. Concerned about the spread of a certain disease the health department launch a long term campaign aimed at increasing the population's awareness of the risks. To monitor the success of the campaign amongst various sections of the community all newly diagnosed sufferers are noted according to a number of characteristics. The graphs below show the percentage of newly diagnosed sufferers in each category, before the campaign started and 5 years into the campaign.



(a) The percentages for each graph add up to more than 100%! Explain.

(b) Would it be correct to say that the graphs indicate that the number of newly diagnosed sufferers who are male has decreased but the number who are under 30 has increased? Justify your answer.