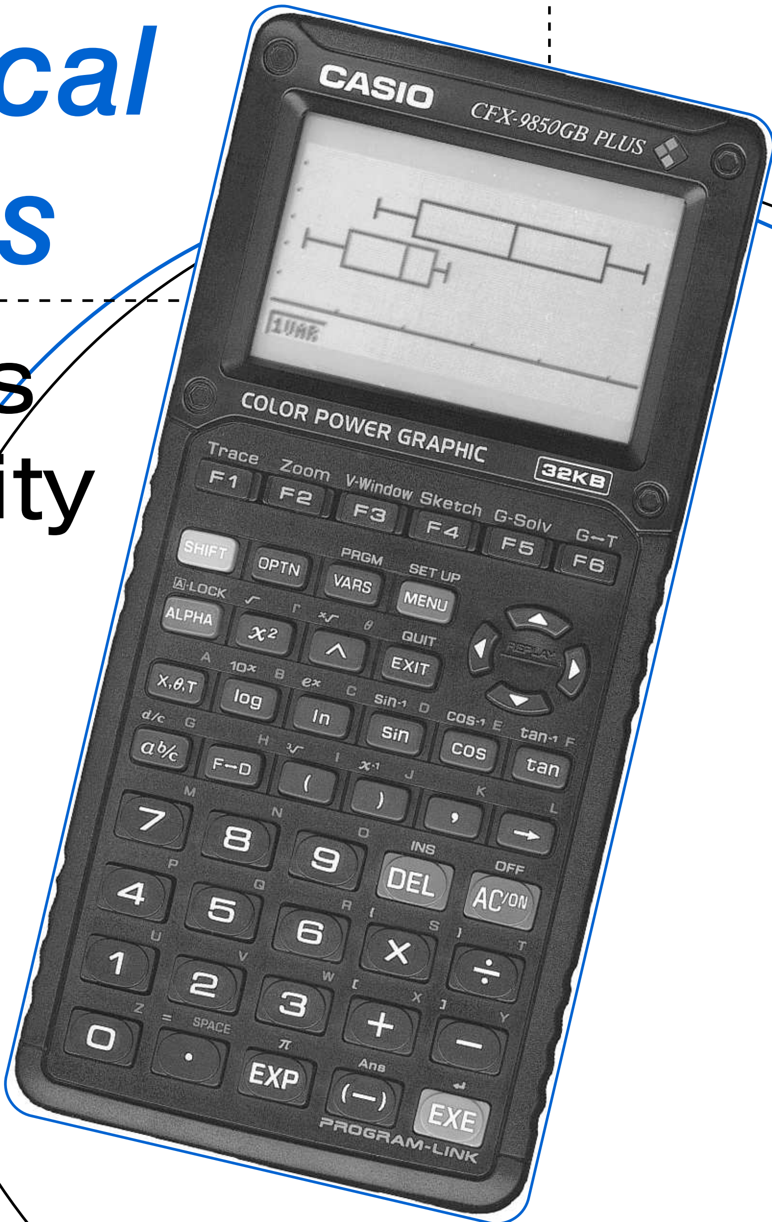
NSW GENERAL MATHEMATICS Preliminary Course

Mathematical Interactions

Data Analysis and Probability



Barry Kissane
Anthony Harradine
Anthony Boys

Mathematical Interactions:
Data Analysis and Probability

Published by Shriro Australia Pty Limited
72-74 Gibbes Street Chatswood, NSW 2067, Australia

A.C.N. 002 386 129

Telephone: 02 9370 9100

Facsmile: 02 9417 6455

Email: casio.edsuppport@shriro.com.au

Internet: http://www.school.casio.com.au

Typeset by Peter Poturaj of
Haese & Harris Publications

ISBN 1 876543 24 8

About this book

Calculators are too often regarded as devices to produce answers to numerical questions. However, a graphics calculator like the Casio CFX-9850GB PLUS is much more than a tool for producing answers. It is a tool for exploring mathematical ideas. We have written this book to offer some suggestions of how to make good use of this tool for the particular areas of data analysis and probability.

We assume that you will read this book with the calculator by your side, and make use of it as you read. Unlike some mathematics books, in which there are many exercises of various kinds to complete, this one contains only a few ‘interactions’ and even fewer ‘investigations’. The learning journey that we have in mind for this book assumes that you will complete *all* the interactions, rather than only some. The investigations will give you a chance to do some exploring of your own.

We also assume that you will work through this book with a companion: someone to compare your observations and thoughts with; someone to help you if you get stuck; someone to talk to about your mathematical journey. Learning mathematics is not meant to be a lonely affair; we expect you to interact with mathematics, your calculator and other people on your journey.

Throughout the book, there are some calculator instructions, written in a different font (like this). These will help you to get started, but we do not regard them as a complete manual, and expect that you will already be a little familiar with the calculator and will also use the *Owner’s Manual*, our *Getting Started* book and other sources to develop your calculator skills.

Data Analysis and Probability are two of the topics in General Mathematics. Data analysis can involve data collected by other people or data that you collect for yourself. Both of these aspects are explored in this book. Meaningful interpretation of data analysis also requires some understanding of probability, which is why we have included it in this book. We have also included some exciting material on analysing data produced through simulation. An important source of data, understanding simulation also requires an understanding of the basic ideas of probability. Although we have sampled some of the possible ways of using a graphics calculator to learn about these topics, we have certainly not dealt with all of them.

When this symbol appears in the book, you are being reminded that the data or calculator programs that you will need to use are available in electronic form on the Internet at http://www.school.casio.com.au. Consult our *Getting Started* book to learn how to use the FA-122 Program Link to transfer data and programs into your calculator. This will save you some laborious key work.

We hope that you enjoy your journey!

Barry Kissane

Anthony Harradine

Anthony Boys

Table of contents

Drawing and using histograms

Every day we are faced with problems that we would dearly like to solve or situations that we would love to change for the better.

In the course of investigating such problems we often need to

- collect information (data),
- process this data in some manner and
- make conjectures about the situation or propose possible solutions to the problem that are based on findings from the processing of the information.

It is critical that our conjectures and solutions are based on the statistical facts that the processing of the data reveals. If they are not they will be no more than simple guesses. In this book you will learn about concepts and tools involved in the investigation of situations involving *continuous* data. The first step in such investigations is to produce a graphic display that illustrates the variation that the data exhibits. A *histogram* is a suitable tool to use for this purpose.

We have all been in a car that has travelled through a road work zone. Such zones often have a speed limit considerably lower than that of the stretch of road immediately before the road work zone to ensure the safety of the workers.

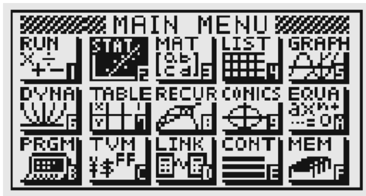
Workers regularly complain that drivers either do not bother to slow down or do not slow down enough. Miles, a year 11 student, decided to investigate these complaints. He obtained a speed gun and visited one road work zone in the suburb of Collinswood where the speed limit was 25 km/hr. The speed limit of the stretch of road immediately before this was 60 km/hr. Miles recorded, to the nearest km/hr, the speeds of the first 100 cars that passed through the zone. The data are given below:

40	44	48	29	28	32	26	26	38	21
31	46	32	37	52	39	28	25	26	41
19	30	54	55	44	33	28	32	30	34
29	53	29	34	28	38	34	49	24	36
26	24	28	35	22	37	36	58	49	21
42	40	48	43	21	25	32	31	25	31
25	33	28	42	34	29	40	24	33	33
37	32	44	31	30	21	31	24	53	19
53	34	26	27	58	29	35	26	45	32
26	56	25	52	38	24	36	24	31	36

Interaction A

1. What is the variable about which Miles collected data?
2. The drivers who had their speeds recorded can be considered as a sample of a population. Identify the population.
3. Would this sample be considered as a random sample? Explain.

Once Miles had collected these data he began processing this data to see what mysteries it had to reveal. Follow the instructions below to produce a frequency histogram to illustrate Miles' data:

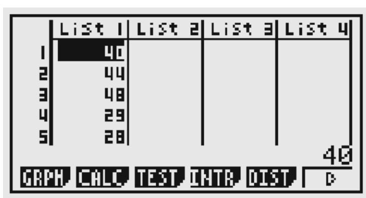


With the calculator turned on and the main menu visible, use the arrow key to highlight the STAT menu (shown opposite).

Then press the blue EXE key (alternatively, simply press 2). This screen will result.

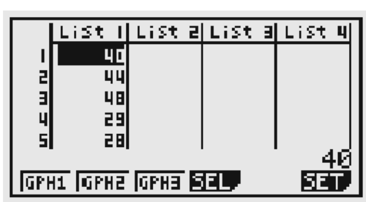
Now access SET UP (SHIFT then MENU) to arrive at the preferences for this module. Be sure each option is set as shown below. Use the down arrow key to scroll down to the lower options.

Press the EXIT key to return to the window containing the lists.



To enter data into the lists, simply use the number key pad and press the blue EXE key after each entry. Enter Miles' data on the speed of the cars into List 1. If you make an error in typing before you press the EXE key, use the back arrow key to repair it. If you have pressed the EXE key, simply use the up arrow to select the wrong data point and re-enter it.

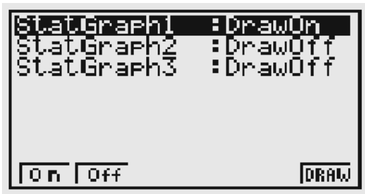
The title for each list cannot be changed, so you will need to document that List 1 contains Miles' data.



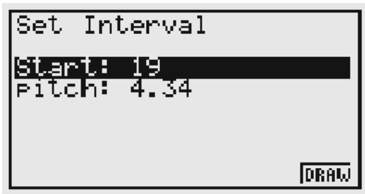
Use GRPH (F1) to reveal the options shown opposite. We are able to set up three different graphs.

Use SET (F6) to set up from one to three graphs. You can select which graph you wish to set up by using GPH1 (F1) or GPH2 (F2) or GPH3 (F3). Graph 1 should be selected now. Use the down arrow key and the F keys to set the options as shown opposite.

Press the EXIT key.

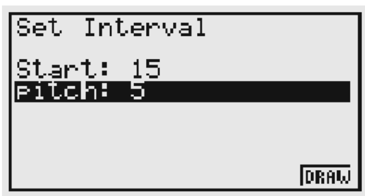


Access SEL (F4) to select which of the three graphs you want the calculator to draw. All three graphs can be displayed if we wish; they need not be the same type of graph. Turn StatGraph 1 on and the other two off.



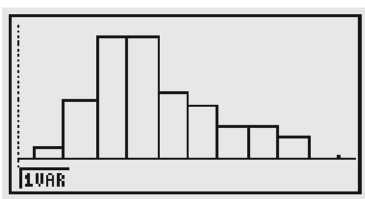
Activate DRAW (F6) and the screen opposite will appear.

The calculator has suggested the first lower group limit (Start) and the group width (pitch) for us. We can change this to whatever we want. A quick scan of the data shows that the lowest speed was 19 km/hr and the highest speed was 58 km/hr.



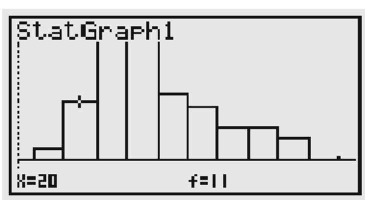
Use the arrow keys to select the Start and the pitch and change them to 15 and 5 respectively.

You should always choose these yourself to ensure the histogram shows the data grouped in whatever way you desire.



Activate DRAW (F6) and the histogram opposite will result. Note the 1VAR option at the bottom left of the screen. We will want to use this later.

Use TRCE (SHIFT then F1) to ‘arrow’ across the tops of the frequency histogram’s bars to determine the frequency values for each group (denoted opposite by f=11).



The X value is the lower group limit, even though the calculator displays it in the centre of the bar.

Do not delete this data from List 1 in your calculator. You can choose to remove the title StatGraph1 on screen by accessing SET UP and turning GraphFunc off.

Interaction B

1. Arrow around to get a feel of the frequency of each group. Use this histogram to produce a frequency distribution table.
2. What is the shape of the distribution illustrated by this histogram?
3. Miles was asked by the *Roads and Traffic Authority* to explain what the shape of this distribution revealed about the speeds of the drivers (in an overall sense) that he had measured. What should have Miles said?
4. Recall that the speed limit in this zone was 25 km/hr. Use your frequency distribution table to determine how many drivers travelled below the speed limit of 25 km/hr. How many travelled at the limit? How many travelled above the limit?
5. Is the shape of a distribution alone adequate to describe what is going on in this situation? What other information would you like to find out about this distribution to help you form a conjecture about how drivers behave in 25 km/hr road work zones?

Miles extended his study to investigate the speed at which drivers travelled through a 25 km/hr road work zone when the stretch of road before the 25 km/hr zone had a speed limit of 100 km/hr. He repeated the procedure outlined earlier in such a location in the suburb called Stirling. The speeds of the 100 cars that he sampled at this location are given below:

35	46	47	54	45	43	25	36	51	37
55	53	44	54	44	28	36	32	47	29
35	47	29	47	35	29	53	31	38	39
37	26	49	38	32	33	35	37	51	37
37	42	33	35	50	44	55	51	45	42
37	29	37	45	37	30	36	42	25	33
53	45	45	49	31	52	44	50	37	39
52	41	41	50	49	43	38	38	39	51
35	27	35	26	61	32	43	34	30	44
52	34	31	53	39	50	40	42	31	49

Transfer or enter these data into List 2 in your calculator. Remember these data are available on the disk that accompanies this book and are also available on the Internet at http://www.school.casio.com.au.

Interaction C

1. Produce a histogram to illustrate the speed distribution for this situation using `StatGraph` 2. Make this histogram orange.
2. Form a frequency distribution table from this distribution.
3. What is the shape of the distribution illustrated by this histogram?
4. Miles was asked by the *Roads and Traffic Authority* to explain what the shape of this distribution revealed about the speeds of the drivers (in an overall sense) that he had measured in this situation. What should have Miles said?

5. Use your frequency distribution table to determine how many drivers travelled below the speed limit of 25 km/hr. How many travelled at the limit? How many travelled above the limit?
6. Set up your calculator so that `StatGraph 1` and `2` are turned on. (Use the `SEL` option). This will cause the calculator to draw the histograms of both situations on top of each other. What does this graphical representation reveal about the difference in driving behaviour (with respect to speed reduction) of the drivers sampled who travelled from a 60 km/hr to 25 km/hr area compared to those who went from a 100 km/hr to 25 km/hr area? Use the `TRACE` option (or frequency distribution tables) to gain further evidence to support your opinion - arrow up and down to interchange between the histograms.
7. Do you think that Miles will be able to form a conjecture about the behaviour (with respect to speed reduction) of Australia drivers when faced with the two situations you have studied? Give reasons to support your opinion.

Investigation

Use the data that you have entered into your calculator to investigate the effect on the histogram produced if you use different pitch values. Try both small and large values for the pitch and comment about the shape and general usefulness of the resulting histograms.

Measuring and using the centre of a distribution

Stephanie, also a year 11 student, was intrigued that she could purchase differently priced spools of fishing line that are rated with the same breaking strain. She decided to investigate this situation, wondering if the cheapest brand actually measured up to its quoted breaking strain.

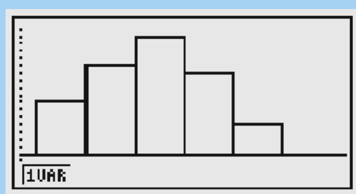
Stephanie purchased five spools of the cheapest brand of fishing line she could find with a breaking strain of 1.8 kg. She then cut each spool into 1 metre lengths and randomly selected ten 1 metre lengths from each spool. On a homemade breaking strain tester, she then recorded the weight required to break each piece of line. This process was repeated for each of the five spools purchased. The breaking strain data (in kilograms) Stephanie collected is given below.

1.42	1.45	0.95	1.00	1.05	1.05	1.10	1.15	1.15	1.15
1.30	1.30	1.35	1.35	1.50	1.45	0.70	0.85	0.80	0.85
1.10	0.90	1.00	1.10	1.20	1.10	1.05	1.20	1.30	0.85
0.80	0.70	1.00	1.00	0.95	0.90	1.00	1.25	1.10	1.10
1.40	0.85	0.70	0.80	0.70	0.60	1.20	0.95	0.70	1.20

Transfer or enter the above data into List 3 of your calculator.

Interaction D

1. Produce a histogram using `StatGraph 1` (`Start: 0.6, pitch: 0.2`) that illustrates the breaking strain distribution of this sample of data. Make this histogram blue. Be sure that it looks like the one shown opposite. Use the `TRACE` option to determine the frequencies of each group and then draw a fully scaled and fully labelled histogram of the breaking strain data in your book.

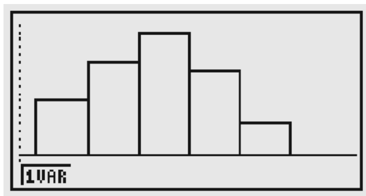


2. State the shape of this distribution and hence describe how the breaking strains of the individual pieces vary.

We should never have expected that the fifty pieces of line would have identical breaking strains. Production processes are open to errors and we find that all quantities like breaking strain carry with them some amount of variation.

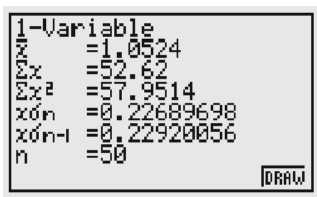
So we should be happy with all pieces not registering 1.8 kg. However, if we were to be happy that the quoted breaking strain of the line was 1.8 kg, we would have expected that about half of the pieces broke with a weight below 1.8 kg and about half above 1.8 kg. In other words:

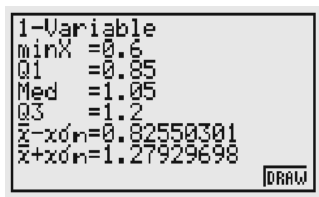
We would hope that 1.8 kg (or about that) would be the centre of the distribution.

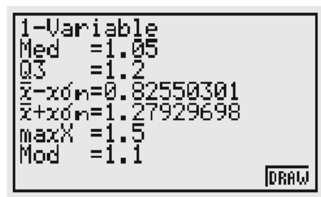


You should know that there are two commonly used statistics that measure the centre of a continuous distribution. They are the mean and the median. These values can be quickly determined using the calculator. Recall the histogram that we produced for these data. Produce this on your calculator again. Note the 1VAR option at the bottom left of the screen.

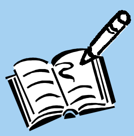
Activate 1VAR (F1) and the calculator will calculate the median and the mean of this data, among other things. You will need to use the down arrow to see all the values on offer – see below.







In this situation the median and the mean are very close together – as we would expect with a symmetrical distribution.



Interaction E

1. Explain how the fact that the median is approximately 1 kg helps you to begin to argue that this brand of line does not have the breaking strain that it is advertised as having.
2. What other feature(s) of this distribution should you consider when building your argument? Explain your answer.
3. What assumptions would you be making if you were to make the conjecture that the breaking strain was more like 1 kg than 1.8 kg?

Stephanie continued her investigation. She purchased five spools of a 'mid-range' priced 1.8 kg line. She repeated the same process on these spools as she used for the cheapest brand. The data she collected are given below.

1.05	1.55	1.60	1.70	1.75	1.85	1.85	1.95	1.95	1.95
2.00	2.05	2.05	2.20	2.25	2.35	0.85	1.20	1.55	1.50
1.90	1.85	1.85	1.80	1.80	2.05	1.95	2.10	1.90	2.20
2.15	2.40	0.75	1.25	1.50	1.40	1.85	1.80	1.90	1.95
1.65	2.05	1.90	1.90	2.00	0.95	1.95	1.85	2.15	1.80

Transfer or enter these data into List 4 in your calculator.

Interaction F

1. Produce a frequency histogram to illustrate the breaking strain distribution for the mid-range priced line using `StatGraph` 2. Make this histogram orange when you set up `StatGraph` 2.
2. Form a frequency distribution table from this distribution.
3. What is the shape of the distribution illustrated by this histogram?
4. Determine the mean and the median of this distribution.
5. Set up your calculator so that `StatGraph` 1 and 2 are turned on (use the SEL option). Draw the histograms of both situations on top of each other. What does this graphical representation reveal about the relative breaking strain of the cheapest line compared to the 'mid-range' line.
6. What conjecture(s) do you feel Stephanie could make about the actual breaking strain of the 'mid-range' priced line. Why?
7. What other information about the distribution would give your conjecture(s) more support?

Interaction G

1. You should still have Miles' data in List 1 and 2 of your calculator. Determine the mean and median for each data set. Record these statistics in a table, along with the shape of each distribution.
2. Explain the similarity or difference between the mean and median of each distribution by making reference to the shape of each distribution.

Measuring and using the spread of a distribution

A new drug was developed that is claimed to lower the cholesterol level in humans. A heart specialist was interested to know if the claim made by the company selling the drug was accurate. He enlisted the help of 50 of his patients. They agreed to take part in an experiment in which 25 of them would be randomly allocated to a group that would take the new drug and the other 25 would take an identical looking pill that was actually a placebo (a sugar pill that will have no effect at all).

The random allocation was carried out by the statistician who was to analyse the data collected in the experiment, so neither the patients nor the doctor researching this situation knew who was talking the drug and who was talking the placebo.

All participants had their cholesterol level measured before starting the course of pills and then at the end of two months of taking the pills, they had their cholesterol level measured again. The data collected by the doctor are given below.

CHOLESTEROL LEVELS OF ALL PARTICIPANTS BEFORE THE EXPERIMENT

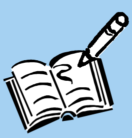
7.1	8.2	8.4	6.5	6.5	7.6	7.2	7.1	6.1	6.0
8.5	5.0	6.7	7.3	8.9	6.2	6.3	7.1	8.4	7.4
7.6	7.5	6.6	8.1	6.2	7.0	8.1	8.4	6.4	7.6
8.6	7.5	7.9	6.2	6.8	7.5	6.0	5.0	8.3	7.9
6.7	7.3	6.0	7.4	7.4	8.6	6.5	7.6	6.3	6.2

CHOLESTEROL LEVELS OF THE 25 PARTICIPANTS WHO TOOK THE DRUG

4.8	5.6	4.7	4.2	4.8	4.6	4.8	5.2	4.8	5.0
4.7	5.1	4.4	4.7	4.9	6.2	4.7	4.7	4.4	5.6
3.2	4.4	4.6	5.2	4.7					

CHOLESTEROL LEVELS OF THE 25 PARTICIPANTS WHO TOOK THE PLACEBO

8.4	8.8	6.1	6.6	7.6	6.5	7.9	6.2	6.8	7.2
7.5	6.0	5.7	8.3	7.9	6.7	7.3	6.1	7.4	8.4
6.6	6.5	7.6	6.1	8.2					



Interaction H

The experiment is said to be a 'double blind' experiment. What is meant by this?

Three lists are needed to enter this data into the calculator. You should have data in Lists 1 to 4 at present and so only two of the lists in your calculator should be free. So that you do not have to re-enter all of the data from the earlier situations again, use a new `File`. The file you are working in now is `File 1`.

The calculator has six files that each contain six lists. These can not be named on the calculator so it is a good idea to keep a record of what is in them on paper. Write down the investigation and variable for which the data in each list of `File 1` applies.

From the `Main Menu`, select the `List (4)` module.

Access `SET UP` (`SHIFT` then `MENU`). The window opposite should be seen.

Activate `File2 (F2)` and then press `EXIT`.

An empty list set should be available.

You have not deleted the other data; they are just hiding in `File 1`.

Now enter or transfer the ‘all before’ cholesterol data into `List 1`, the ‘drug’ data into `List 2` and the ‘placebo data’ into `List 3`.

When asked what information she wanted from the analysis of these data the doctor said,

As well as being interested in the centre of the distributions, I am also interested in the values in between which the middle 50% of cholesterol levels fell when the data are ranked in ascending order and hence how spread the middle 50% of scores are.

The doctor is wanting to know something about the *spread* of the data.

A commonly used set of summary statistics that help in cases like this is called the *five number summary*. This includes the following statistics:

Minimum score – Min
First quartile (the quarter marker) – Q1
Median
Third quartile (the three quarter marker) – Q3
Maximum score – Max

A useful statistic that is derived from the first and third quartiles is the *interquartile range* (IQR). It gives the range of the middle 50% of the scores when ranked in ascending order. To calculate the IQR, subtract the first quartile from the third quartile.

$$\text{IQR} = \text{Q3} - \text{Q1}$$

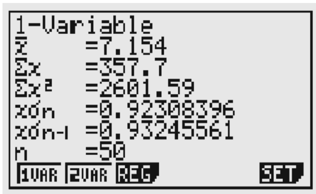
The range is another measure of spread used sometimes. It is calculated by subtracting the lowest score from the highest score.

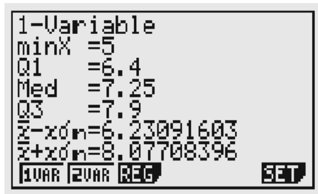
$$\text{Range} = \text{Max} - \text{Min}$$

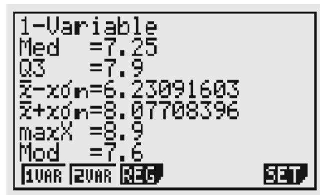
These statistics can be efficiently calculated using the calculator. We shall now do this for the ‘all participants before’ data.

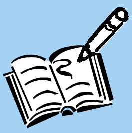
Press MENU and then press 2 to enter the STAT module. Activate CALC (F2) and then use SET (F6) to set up the calculations you want. Set the options as shown on the screen opposite. Setting the 1VAR XList to List 1 means that statistics for List 1 will be calculated.

Press EXIT and then use 1VAR (F1) to make the calculator display the summary statistics. You should achieve the results seen below. Use your arrow keys to scroll down and reveal all that has been calculated. Much of this is not yet required.







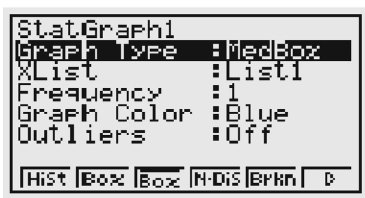


Interaction 1

Use the calculator to determine the five number summaries and calculate the interquartile ranges for each group of subjects. Record these statistics in an appropriate table.

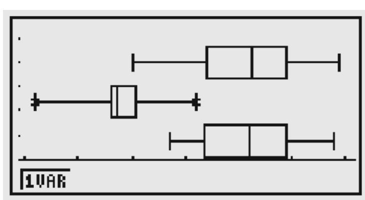
The five number summary can be displayed graphically using a boxplot. This can be a very useful graphic when comparing two distributions or comparing a distribution to a standard value.

We can produce up to three boxplots on the calculator at any one time. We will now produce boxplots for the three data sets.



Enter STAT mode and activate GRPH. Use SET to enter the screen that will allow you to set up the type of graph you wish to use. You can set up all three graphs by choosing GPH1, GPH2 or GPH3. Use the down arrow key to select Graph Type. Use the continuation key (F6) to find more options and then activate BOX (F2). Set each option as shown opposite.

Use the arrow keys to select StatGraph 1 and then set up StatGraph 2 to be a MedBox of List 2 and to be orange in colour. Set up StatGraph 3 to be a MedBox of List 3 and to be green in colour. Press EXIT when you are finished.



Use SEL to turn on each of the possible StatGraphs and then use Draw (F6) to instruct the calculator to draw the three boxplots.

You can use TRCE (SHIFT then F1) to determine the values of the five number summary from these boxplots. Try it out.

This graphic shows clearly that:

> the middle 50% of the before values and the placebo values have very similar ranges (approximately 6.4 to 7.9) whereas the middle 50% of values of the drug group were spread over a much smaller range and are generally much lower than the other two groups values (4.6 to 5.05).

Interaction J

1. Does the analysis you have carried out suggest that the drug has had an effect on cholesterol level? If so, what is the effect? Support your answer with statistical evidence.
2. Return to the speed limit investigation. You should have these data still entered in File 1 of your calculator. Determine the five number summaries and produce boxplots for the two data sets. Write a paragraph in which you compare the middle 50% of speeds in each group to the speed limit.
3. Return to the fishing line investigation. You should have these data still entered in File 1 of your calculator. Determine the five number summaries and produce boxplots for the two data sets. Write a paragraph in which you compare the middle 50% of weights at which the samples from each brand broke, to the advertised breaking strain of each brand.
4. Experiment with the other box option that is available. How are the boxplots produced by this command different from those you have been producing?

Investigation

The data collected by the doctor during the cholesterol test can be analysed in a different manner. There were two groups of people in this investigation – those who took the placebo and those who took the drug. We could create another variable to consider, namely 'cholesterol value before treatment (Cb) – cholesterol value after treatment (Ca)'. Complete the table on page 17.

Transfer or enter this data into the list in `File 3` of your calculator. Investigate the variable Cb−Ca with the aim of collecting evidence to support the conjecture that the drug will have an effect on the cholesterol level of those who take it. (Hint: What would the value of Cb−Ca be if the drug makes no difference?)

Write a short report in which you use the statistical evidence you have collected to form an argument that allows you to conclude with the conjecture given above.

Placebo Group				Drug Group			
Subject Number	Cb	Ca	Cb − Ca	Subject Number	Cb	Ca	Cb − Ca
1	7.1	8.4		26	7.0	4.8	
2	8.2	8.8		27	8.1	5.6	
3	8.4	6.1		28	8.4	4.7	
4	6.5	6.6		29	6.4	4.2	
5	6.5	7.6		30	7.6	4.8	
6	7.6	6.5		31	8.6	4.6	
7	7.2	7.9		32	7.5	4.8	
8	7.1	6.2		33	7.9	5.2	
9	6.1	6.8		34	6.2	4.8	
10	6.0	7.2		35	6.8	5.0	
11	8.5	7.5		36	7.5	4.7	
12	5.0	6.0		37	6.0	5.1	
13	6.7	5.7		38	5.0	4.4	
14	7.3	8.3		39	8.3	4.7	
15	8.9	7.9		40	7.9	4.9	
16	6.2	6.7		41	6.7	6.2	
17	6.3	7.3		42	7.3	4.7	
18	7.1	6.1		43	6.0	4.7	
19	8.4	7.4		44	7.4	4.4	
20	7.4	8.4		45	7.4	5.6	
21	7.6	6.6		46	8.6	3.2	
22	7.5	6.5		47	6.5	4.4	
23	6.6	7.6		48	7.6	4.6	
24	8.1	6.1		49	6.3	5.2	
25	6.2	8.2		50	6.2	4.7	

Comparing relative frequency and probability

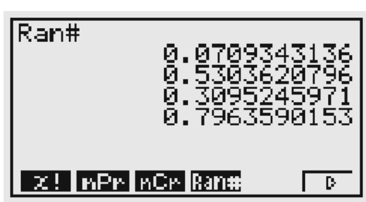
The mathematics of probability is concerned with things that happen at random, or unpredictably. A common example is a (fair) raffle in which each ticket has an equally likely chance of winning. Nobody knows which ticket will win until the raffle is drawn. In the same way, when you toss a coin in the air, you can't predict reliably whether it will fall 'heads' or 'tails'. Most things in the world have some random aspects in them, which is why it is so hard to predict the future accurately.

If you get a lot of information about a random event, it is often possible to see some patterns, which will help to understand it better. One way to see this uses a standard six-sided die.

Interaction K

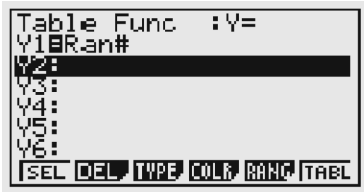
1. Toss the die ten times, recording the number on the uppermost face each time. Which number(s) occurred most often? Check with other students.
2. Toss the die another ten times. Can you see any patterns in the numbers that are thrown? For example, how often are two numbers in succession the same?
3. How many times has each number been thrown? Compare your answers with other people.
4. Each person in your class has tossed a die 20 times. Work out how many fives were thrown altogether. Check how many of the other numbers were thrown altogether. Write a sentence or two to explain what you notice.

You can use your calculator to generate numbers at random. Enter RUN mode (by pressing MENU and 1). To get the random number command, press OPTN and then access the PROB menu (F6 followed by F3). Then Ran# (F4) is a command to give a random number larger than zero and less than 1 each time it is used.



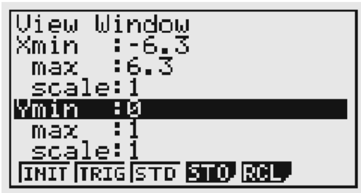
In the screen at right, EXE was pressed four times after Ran# was entered, so that four random numbers were generated.

To see a lot of random numbers, it is easier to make a table than to press EXE many times. Enter TABLE mode (by pressing MENU and 7). Enter the Ran# command in the function list and use RANG (F5) to specify a table of 100 values, as shown below.



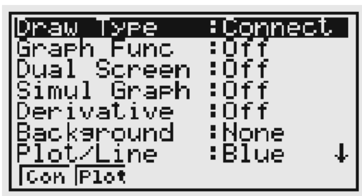
Press EXIT and then activate TABL (F6) to generate a table of 100 random numbers, which you can explore with the cursor keys.

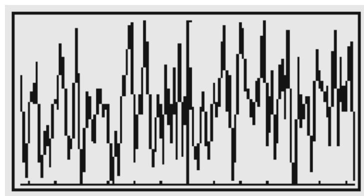
You can also get a picture of randomness by drawing a graph of the Ran# function. First enter GRAPH mode by pressing MENU and 5. The Ran# function will already be in the function list.



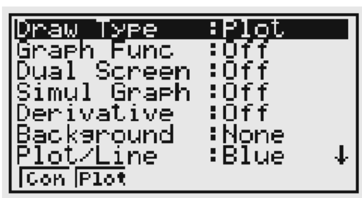
Set up the viewing window so that the y values are between 0 and 1, as shown here.

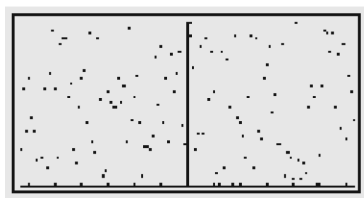
Your graph will look a bit different from ours. To get a new graph, press EXIT after a graph is finished and then DRAW (F6) again.

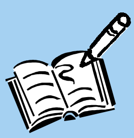




Notice that the graph above is drawn with the points connected, which does not make much sense here; the graph below is drawn with the points plotted (not connected). You can change from one type of graph to the other by first accessing SET UP (SHIFT MENU) and then adjusting the Draw Type.







Interaction L

1. In RUN mode, use Ran# to generate six random numbers. How many are larger than 0.5? Repeat this experiment a few times. Compare your results with others.
2. Make some tables of the Ran# function. Do this a few times. Write a paragraph to describe the numbers generated.
3. Draw some graphs of the Ran# function. Notice that each one is different. (Check this by looking carefully at one part of the graph each time – like

the top right corner.) How would you describe the sets of random numbers produced?

4. The Ran# function should produce random numbers evenly spread between 0 and 1. For example, about half the time the result should be larger than 0.5. Does this seem to be the case for your tables and graphs?

The *probability* of an event is a theoretical idea that is useful when there are a number of equally likely outcomes. A good example is the tossing of a standard six-sided die. If we assume that the six possible outcomes (1, 2, 3, 4, 5 and 6) are all equally likely, then we can find the probability of events based on these.

For example, the probability of getting a four on one toss of the die is 1/6. Similarly the probability of tossing an odd number on one toss of the die is 3/6 or 1/2 (since there are three favourable outcomes and six possible outcomes).

The *relative frequency* of an event is a practical idea: it describes what fraction of the time an event *actually* occurred. In the long run, the relative frequency of an event will be close to the (theoretical) probability.

For example, if you toss a die 100 times, you should expect to get a four about one sixth of the time; that is, around 16 or 17 times. But the relative frequency describes how many times you actually got a four. If you tossed a four 18 times out of the 100 tosses, the relative frequency would be 18/100 or 18% or 0.18.

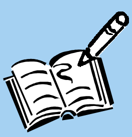
You can use your calculator to *simulate* events with a certain probability and see what the relative frequencies of the events are. This is a kind of ‘experimental probability’.

For example, the command `Int (Ran# + 0.3)` will give the value of 1 (a “success”) 30% of the time and a value of 0 (a “failure”) 70% of the time. That is, it will simulate an event with a probability of 0.3 or 30%. In RUN mode, the `Int` command is accessed by first pressing `OPTN` and then `F6` followed by `NUM` (`F4`).

To simulate events with other probabilities, just change the numerical value. For example, to simulate the event of getting a four when tossing a die, use `Int (Ran# + 1/6)`. To simulate getting a head when tossing a fair coin, use `Int(Ran# + 1/2)`. Use the fraction key to enter fractions like 1/6 and 1/2.

This screen shows the calculator being used in RUN mode to simulate five coin tosses. Once the command was entered, the EXE key was pressed five times. Each press simulated another toss.

In this sample, there were two heads out of the five simulated tosses, so that the relative frequency of heads was 2/5 or 40%.



Interaction M

1. Simulate five coin tosses. Write down the relative frequency of heads. Repeat this a few times. (Each time, press the left cursor key and then press EXE five times.) Does the relative frequency ever reach 100%? Is it ever 0%? Compare with your partner.
2. People often think that if something has a probability of 1/2, then it is likely to happen half the time. To test this, use your calculator to simulate six coin tosses. Do this several times and describe your results.

3. Simulate an event with probability one sixth, such as tossing a four on a standard die. How many tosses does it take until you get a four?
4. A company claims that 90% of its TV sets don't need to be repaired under warranty. That is, they claim that the probability that a TV set lasts for the warranty period is 90%. Use your calculator to simulate a series of events with probability 90%; how long does it take until you get a failure? Compare your results with other people.

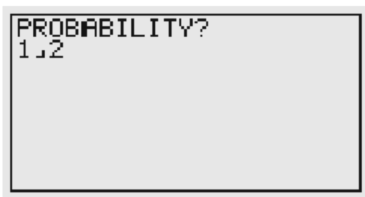
When it is too hard to work out a theoretical probability, people often use the relative frequency for a large number of cases as an estimate.

For example, the company mentioned in Interaction M.4 may have found that 9000 out of 10 000 TV sets lasted the warranty period, and used the relative frequency of 9000/10 000 or 90% as an estimate of the probability.

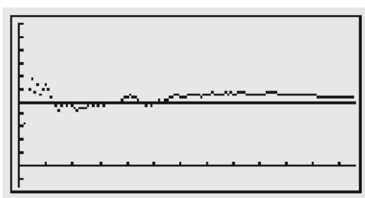
You can use your calculator to see what happens in the long run. The calculator program called `LONGRUN` will allow you to simulate a sequence of events with a certain probability and to find the relative frequency of the event.

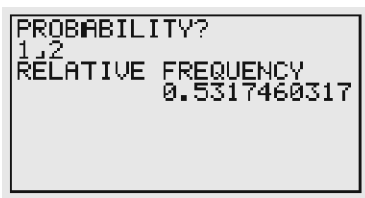
Transfer the program `LONGRUN` to your calculator.

To use the program after it has been entered into your calculator, enter the Program mode by selecting `PRGM` from the main menu or just by pressing `MENU` and then the `log` key. Highlight the program name (`LONGRUN`) and press `EXE`.



The program will ask you for a probability. Enter the desired value and press `EXE`. In this screen, a probability of 1/2 has been entered.





The calculator will then simulate trials with a successful outcome having probability 1/2 and draw a graph of the relative frequency after each simulation.

In this case, notice that at first the relative frequency was larger than 1/2 (shown by the horizontal line), but that after about fifteen simulations it was a little under 1/2.

By the end of the simulations (127 of them) the relative frequency is a little over 1/2.

In fact, the screen above shows that the relative frequency in this case was approximately 0.532, or 53.2%, which is quite close to the probability of 1/2.

When the program has finished, you can use `G-T` (`SHIFT F6`) to switch between the graph and the numerical result. On the graph, the vertical scale shows a tick mark at every 0.1, while the horizontal scale shows a tick mark at every ten simulations.

Interaction N

1. Use program LONGRUN to simulate tossing a fair coin 127 times. You should expect the relative frequency to be close to 0.5. Why can it not be exactly 0.5?
2. The probability of getting a four on a die is 1/6. Use the program to simulate 127 dice tosses. How close is the relative frequency to the theoretical probability? Compare your answer with others.
3. Use the program to simulate the production of the TV sets described above, which have a probability of 0.9 of lasting to the end of the warranty period. Do this several times. How likely does it seem that the relative frequency will drop below 85% for 127 TV sets? Discuss your conclusions with your partner.
4. Of course, program LONGRUN makes no sense for ‘probabilities’ that are less than 0 or greater than 1. Explain what happens when a probability of 0 or a probability of 1 is used to simulate events.

Investigation

When someone answers multiple choice questions with four choices, they have a probability of 1/4 of guessing the correct answer (with a completely random guess). Investigate how likely it is that this sort of guessing strategy will allow someone to get more than half the questions correct on a test.

Experimenting with random samples

Samantha was annoyed because the batteries in her portable CD player had gone flat again. She decided to try and find out how long a new set of batteries would last.

She knew that the life of the batteries would vary a bit, but wanted to get a good idea of how long to expect them to last. She paid careful attention to how many hours each of the next five sets of batteries lasted. To the nearest hour, she found they lasted 41, 39, 37, 42 and 37 hours.

Interaction 0

1. Give some reasons for the sets of batteries lasting different amounts of time.
2. Find the mean number of hours the sets lasted.

This gave Samantha some information about the battery life. They seemed to last somewhere between 37 and 42 hours. But she wondered whether the results would be similar next time she put in fresh batteries.

Samantha had a *sample* of five battery lifetimes.

In data analysis, we are often interested in some aspects of a *population*, but have to rely on a sample. In this unit, we will see how and why sampling is effective for finding out about a population.

There are many kinds of samples; the most important are *random* samples, for which each member of the population has the same chance of being chosen. In this unit, we will only be studying random samples.

Suppose we had a population of battery lives and wanted to know the mean battery life. We will study how close the means of samples are to the population mean.

(Of course, if we *really* had a population, it wouldn't be necessary to take a sample! This unit will help you to see what happens with sampling, but it does not show what you would usually do. We have to rely on samples because it is usually not possible to study the whole population directly.)

The following tables show a population of 216 battery lives for the CD player.

Notice that there are six tables, numbered 1 to 6, with the rows and columns of each table also labelled 1 to 6.

You can take a random sample from this population using a standard die (i.e. each face of the die has the same chance of being thrown).

Here's how:

1. First throw the die to choose one of the six tables.
2. Then throw the die again to choose a column of the table (labelled in bold on the horizontal axis).
3. Then throw the die again to choose a row of the table (labelled in bold on the vertical axis).
4. The battery life chosen is shown in the column and row of the table (in italics).

To make sure this is clear, check that throwing 6 then 3 then 2 gives a sampled battery life of 32 hours. (In Table 6, the figure shown in the third column and second row is 32.)

Table 1

6	*34*	*36*	*34*	*34*	*40*	*41*
5	*35*	*40*	*41*	*40*	*36*	*28*
4	*42*	*36*	*45*	*40*	*38*	*33*
3	*42*	*28*	*40*	*30*	*40*	*29*
2	*42*	*42*	*30*	*37*	*31*	*43*
1	*39*	*35*	*38*	*33*	*39*	*39*
	1	**2**	**3**	**4**	**5**	**6**

Table 2

6	*34*	*40*	*44*	*45*	*40*	*38*
5	*41*	*44*	*38*	*38*	*37*	*42*
4	*43*	*43*	*44*	*38*	*37*	*36*
3	*34*	*44*	*27*	*34*	*32*	*45*
2	*36*	*30*	*36*	*41*	*42*	*34*
1	*31*	*31*	*39*	*30*	*35*	*32*
	1	**2**	**3**	**4**	**5**	**6**

Table 3

6	*39*	*45*	*30*	*30*	*41*	*44*
5	*35*	*32*	*40*	*39*	*42*	*40*
4	*36*	*36*	*39*	*39*	*39*	*38*
3	*43*	*37*	*35*	*33*	*37*	*31*
2	*38*	*37*	*35*	*37*	*29*	*38*
1	*43*	*33*	*40*	*43*	*34*	*39*
	1	**2**	**3**	**4**	**5**	**6**

Table 4

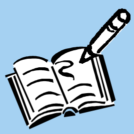
6	*29*	*40*	*42*	*33*	*32*	*29*
5	*42*	*44*	*31*	*35*	*34*	*35*
4	*37*	*41*	*46*	*31*	*41*	*43*
3	*34*	*38*	*43*	*41*	*42*	*33*
2	*42*	*39*	*35*	*41*	*46*	*37*
1	*38*	*38*	*34*	*35*	*37*	*37*
	1	**2**	**3**	**4**	**5**	**6**

Table 5

6	*39*	*40*	*34*	*36*	*40*	*37*
5	*31*	*41*	*38*	*33*	*40*	*39*
4	*36*	*33*	*44*	*38*	*33*	*39*
3	*37*	*32*	*35*	*34*	*29*	*28*
2	*35*	*39*	*41*	*37*	*41*	*31*
1	*35*	*36*	*30*	*38*	*36*	*33*
	1	**2**	**3**	**4**	**5**	**6**

Table 6

6	*37*	*32*	*36*	*34*	*32*	*41*
5	*32*	*32*	*39*	*37*	*43*	*37*
4	*36*	*36*	*41*	*41*	*37*	*38*
3	*39*	*36*	*35*	*37*	*40*	*38*
2	*39*	*33*	*32*	*31*	*35*	*30*
1	*38*	*33*	*36*	*36*	*36*	*34*
	1	**2**	**3**	**4**	**5**	**6**



Interaction P

1. What battery life is sampled by throwing a 5 then a 1 then another 5?
2. Explain why each of the 216 lives has the same chance of being chosen by this method.
3. Use the die to select a random sample of five battery lives. Find the mean battery life of your sample. How does it compare with Samantha's result?
4. Compare your result with those of other students in your class. Write a sentence to describe how much the means of samples of size 5 seem to differ from sample to sample.
5. Consider all the information you have available to you now from the samples you have seen. Estimate what you think the population mean is likely to be.

It is possible, although unlikely, that a random sample might produce the same battery life more than once. (This kind of sampling is called *sampling with replacement*.)

Although it is a perfectly adequate way to select random samples, it is rather tedious to toss a die many times like this. A more efficient way of selecting a random sample is to use a small calculator program for this purpose.

Obtain from your teacher the list containing the population of 216 battery lives and the program SRANSAMP (which stands for *simple random sample*). Transfer the data into `List 1` of a file not yet in use (see page 14) and the program into the program area.

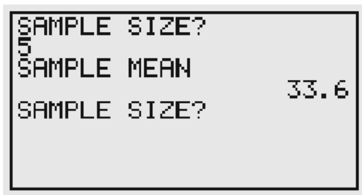
Use the calculator to check that the population mean is approximately 36.96 hours. You can use the program to select a random sample and compare how close the sample mean is to this population mean.

When you run the program, it will ask you to specify the sample size. Enter a whole number and press `EXE`.

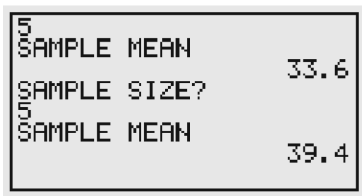
The program will then select a random sample of that size from `List 1` and display the sample mean.

This screen shows an example of this.

This particular sample of size 5 had a mean of 33.6, which is a little smaller than the population mean.



To select another random sample immediately, start by pressing `EXE`. The calculator will again ask you to specify the sample size, as shown here.



Enter the sample size and press `EXE` again. The screen shows an example of this.

For the second sample, the mean was 39.4, which is a little larger than the population mean.

Interaction Q

1. Use program SRANSAMP to select a random sample of size 5. Then enter STAT mode (by pressing MENU 2) and check `List 6`. (You will see the complete sample of five battery lives stored there.)
2. Use the program ten times in succession to select samples of size 5. Each time, write down (on paper) the mean of the sample. When you have finished all ten, enter the means in `List 2` and use your graphics calculator to study them. (E.g., you might draw a box plot or find the mean of the sample means, or do both of these things.)
3. How close to each other are your ten sample means in `List 2`? Compare your results with others. How close are the sample means to the population mean of 36.96?

In general, you will obtain better results from a larger random sample than a smaller one. But of course, it takes longer (and usually takes more money) to select a larger random sample. To see how much better a larger sample is, use your sampling program to experiment.

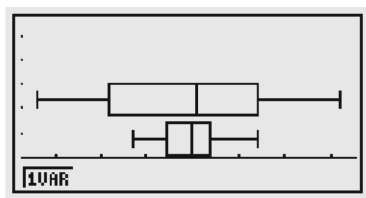
Interaction R

1. Use the program ten times in succession to select samples of size 20. Each time, write down (on paper) the mean of the sample. When you have all ten, enter the means in `List 3` and use your graphics calculator to study the ten sample means. Compare your results with others. How close are the sample means to the population mean of 36.96?
2. How much better is a sample of size 20 than a sample of size 5? Use your calculator to compare the data in `List 2` and `List 3` (e.g. with a box plot of both.) Compare your estimates with others.
3. Now select just one large sample of size 50. Suppose you had only this sample to predict what the population mean is. Compare your prediction with those of other students. How close is your prediction to the actual population mean?

In this case we can compare the actual population mean with the means of the samples used to predict it, but remember that this is usually not the case. By studying how sampling works, provided our sampling is at random, we can make an estimate of the population mean from just one sample (which is usually all we have available). We can also estimate how close our estimate is likely to be! In general, the larger the sample size, the better will be the predictions.

Investigation

When we compared the means of some samples of size 5 with those of some samples of size 20, we obtained the simultaneous box plot shown here.



Which box plot shows which set of sample means? Explain how you can tell.

Choose two other sample sizes, one smaller than 5 the other larger than 20. Repeat the sampling experiment and compare your results with these.

If Samantha were to take a sample of size 30 (i.e. check the lives of 30 sets of batteries and find the mean battery life), how close to the population mean would you expect her sample mean to be?

Analysing simulated data

Some kinds of data are random in nature. A good example occurs in many board games such as* Monopoly *and* Backgammon. *To understand what is happening, it is not possible to gather all the data (the population), because there is an infinite amount of it. (That is, you could play Monopoly all of your life and still not have all the data!). So a sample of data is needed. In some cases, it is possible to use mathematics to theoretically work out what will happen; but until you can do this, analyzing a sample is a good alternative.

You can get a sample of data by tossing a pair of dice many times, but this takes a lot of time. A more efficient method is to *simulate* the data, so that the results are the same, but the method of getting them is different.

The most important kind of sample is a random sample - one for which all the possibilities have the same chance of occurring. Consider the case of rolling a standard die, for which each of the six possibilities: 1, 2, 3, 4, 5 and 6 have the same chance of occurring.

Recall that the calculator's random number command, Ran#, generates random numbers evenly spread between 0 and 1.

To generate suitable data to study a game like *Monopoly* or Backgammon, numbers between 0 and 1 are not directly useful. Instead, you need to transform them to produce counting numbers between 1 and 6. The easiest way to do this is to use the transformation shown in the screen.

Think carefully about the command `Int (6xRan# + 1)`:

Multiplying `Ran#` by six produces numbers between 0 and 6 (instead of just 0 to 1).

Adding 1 transforms the result to give a number between 1 and 7.

The `Int` command takes the integer value of the result, one of {1,2,3,4,5,6}. So the end result of this transformation is always a number between 1 and 6, with each of the six possibilities occurring about as often as each of the others – which is a good model of a standard die.

Interaction 5

1. Try generating several random numbers using the command `Int(6xRan# + 1)`. Check to see that all the results are integers between 1 and 6, just like a standard die.

2. What happens if you use the command `Int (8xRan# +1)` instead of `Int (6xRan# + 1)`? Explain why this happens.
3. How could you use the calculator to generate random tosses of a seven-sided die? Test your answer by trying it out.
4. To model what happens in *Monopoly*, you need to roll a pair of dice not just one. What command will simulate this? Try it and see.

To simulate dice rolls using your calculator, you could make a command like those above and then just press EXE many times.

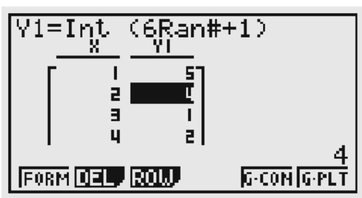
There are two reasons why this is not a good idea, however. In the first place, it will take a long time to gather a reasonable sized sample of data and in the second place, you will need to write down all the results so that you can analyse them. Smart use of the calculator will allow each of these problems to be solved.

The easiest way to get a sample of several dice rolls on the calculator is to generate a table. Enter TABLE mode by pressing MENU and 7 and delete any functions already there.

Insert the command to roll a single six-sided die in the function list, as shown on the screen.

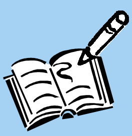
Then use RANGE (F5) to instruct the calculator to generate five values as shown one the screen.

When you activate TABLE (F6), you will see that five separate values are generated, each of them representing a die roll.



The screen shows an example, but of course yours will be different.

Notice that only the first four die rolls are shown; you will need to use the down cursor key to see the other one.

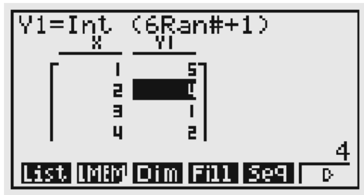


Interaction T

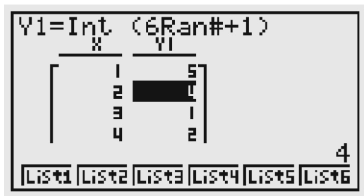
1. Use the calculator to simulate 20 die rolls. Compare your results with those of your partner; e.g. how many times was a six rolled?
2. Return to the function list by pressing EXIT or `FORM(F1)`. Then simulate another set of 20 rolls by pressing TABLE again. Are the results the same as the first time? Explain why.
3. Use the command `Int(6xRan# +1) + Int(6xRan# +1)` to simulate rolling a pair of dice five times and finding the sum of the dice each time. Compare your results with those of your partner. Would you expect the results to be the same or different? Explain why.
4. Use a stopwatch to see how long it takes you to simulate a table of 100 rolls of a pair of dice (after you have made the appropriate adjustments to the RANGE). Compare the time with the time taken to generate 200 rolls of a pair of dice. Explain your results.

As you have seen, it is much quicker to use TABLE mode to simulate a large set of die rolls than it is to do so by repeated pressing of the EXE key in RUN mode. In addition, once you have generated the data, you can examine them using the cursor keys, which you can't do in RUN mode.

It is quite hard to get a good idea of all the data in a large sample such as one with 100 data points. To analyse the data, in order to understand what they can tell us about the results of rolling the pair of dice, it is best to use the calculator to transfer the data into the Statistics area of the calculator.

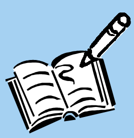


To do this, first move the cursor to the appropriate column of your table of data - the one with the results in it, not the first column which just gives the number of each data point. Press OPTN and then LIST(F1) followed by LMEM(F2) to prepare the calculator to store the column of data into a List Memory.



The screen shows that you have a choice of six data lists, called List1 to List6. (If any of these already contains data, the next step will replace the data with your column of data, so be careful here!) Press the F key associated with a list of your choice to store all the data into that list.

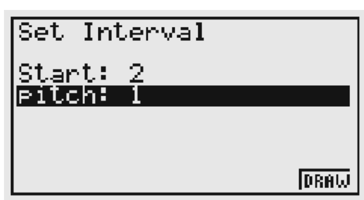
To see the results of this, enter STAT mode by pressing MENU and 2. You will see that all of your data has been automatically transferred, and is now ready for analysis.



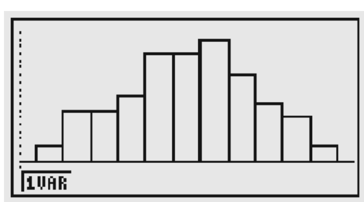
Interaction U

1. Store a set of 100 rolls of a pair of dice into List 1. Write down the first few values in the table. Then check the list in STAT mode to confirm that the first few values are the same and that the list contains 100 values.
2. Return to Table mode and simulate a fresh set of 100 rolls of a pair of dice. Transfer these to List 2.

To see what happens when a pair of dice is rolled, you need to analyse the data. A sample of size 100 is large enough to provide a reasonable idea of the likely results. One way of analyzing the data is to look at a histogram. First check in SET UP that the Stat Wind is set to Auto.



In this case, since only integer values are involved, it's a good idea to set the pitch equal to 1, as shown here.

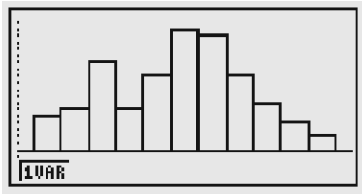


It looks in this case as if there are more values in the middle of the range from 2 to 12 and less values at the ends.

That is, it's a bit more likely that a total of 5 to 9 will be thrown than one of 2 to 4 or 10 to 12.

To see how many there were of each score in your sample, use `Trace` (`SHIFT F1`).

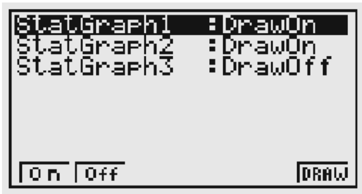
This is only one sample, however. You can use the calculator to check some other samples too.



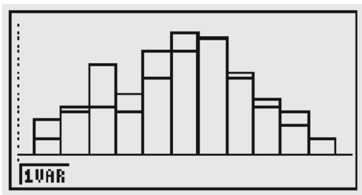
This histogram shows a second sample (which we stored into `List 2` in Interaction U). It's a good idea to use a different colour for this second histogram.

Notice that the results are slightly different, although it is still the case that there are more scores in the middle of the range than at the extremes.

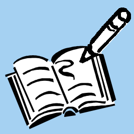
In fact, to compare the two distributions, look at the two histograms on the same screen.



To do this, press `GRPH(F1)` and then `SEL(F4)` to select both `StatGraph1` and `StatGraph2` together, as shown.

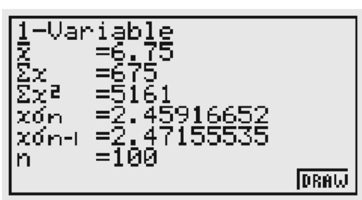


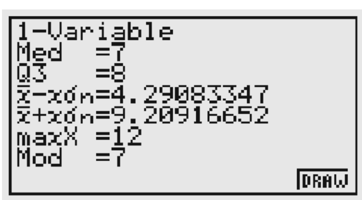
The resulting screen makes it clear that there is a good deal of similarity in the two distributions. This display looks better when the two histograms are drawn in different colours.



Interaction V

1. Draw histograms of your data from Interaction U. How are they like those shown above? How are they different? Trace each histogram to see which total was the most common for each sample.
2. Return to TABLE mode and simulate a fresh set of the sums of 100 rolls of a pair of dice for storing in each of `List 1` and `List 2`. Draw graphs of the two samples and compare them with your first two samples.
3. Use your histograms to write a few sentences that describe the likely results of tossing a pair of two dice and finding the sum of their scores.





Sometimes it is useful to look at numerical summaries of simulated data.

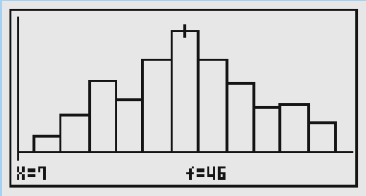
In this case, the mean and median of the samples give a good idea of an 'average' score when a pair of dice is rolled. The mode shows which single result occurred most often. For the first sample shown above, the mean is 6.75, the median is 7 and the mode is also 7.

These three figures, which are quite close to each other in size, are consistent with the graphs showing a distribution that is rather symmetrical about the middle.

The most likely single result is a total score of 7. Why is this? These screens also show that the maximum score was 12 and the minimum was 2, which is also clear from the histogram.

Interaction W

1. Analyse the data in both `List 1` and `List 2`. Compare the results for your two samples. How similar are they to the results shown here? In what important ways (if any) are they different? Compare your results with someone else's.
2. It takes a score of 40 to move a complete revolution of a *Monopoly* board. About how many rolls of a pair of dice do you think this is likely to take, on average?
3. One of the great advantages of using a calculator to study samples of data like this is that it is fairly easy to obtain and analyse a new sample. Return to Table mode, generate a fresh sample, transfer it to List 3 and study it both graphically and numerically.
4. A larger random sample will give you more reliable information than a smaller one. On the `CFX-9850G PLUS`, the largest sample possible has 255 data points. Generate a sample of 255 dice rolls and compare the resulting histogram with the one shown here.



   For example, in this case, a score of seven was obtained 46 times, which is about one time in six. How does this compare with your result?

Investigation

A new board game uses two dice, one of which is a standard die with six sides, while the other has eight sides, numbered 1 to 8.

For each die, all faces are equally likely to occur. What is the most likely result of tossing a pair of dice like this and adding the scores?

Write a brief report, using both graphical and numerical results of studying suitable random samples

Answers

Some of the questions that have been asked do not have a single correct answer. In such cases, MPA (which stands for many possible answers) will be the answer supplied. In many cases some supporting comment is supplied.

Interaction A

1. Speed of cars in a 25 km/hr speed zone. 2. All drivers who travelled through that road work zone in Collinswood.
3. No. A random sample means that each member of the population had an equally likely chance to be a chosen as part of the sample.

Interaction B

1.

Speed (km/h)	15 -	20 -	25 -	30 -	35 -	40 -	45 -	50 -	55 -
frequency f	2	11	24	24	13	10	6	6	4

2. Skewed right.
3. Assuming most drivers traveled at around 60 km/hr before entering the zone, he could say: most people made an effort to slow down but relatively few slowed to, or below, the limit. Only relatively few made little or no effort to slow down.
4. No. Perhaps some ideas of the centre and spread of the distribution.
5. 13 drivers below, 5 at the limit, 82 drivers above.

Interaction C

2.

Speed (km/h)	25 -	30 -	35 -	40 -	45 -	50 -	55 -	60 -
frequency f	10	14	27	15	14	17	2	1

3. Largely uniform, except for the 35 - group and the very small tail at the end.
4. Assuming most drivers traveled at around 100 km/hr before entering the zone, he could say: Drivers seemed to have made an effort to slow down, but none slowed to below the speed limit and only two managed to reach the limit. The driver's speeds were distributed quite uniformly from 25 km/hr to 50 km/hr, except for the modal group of 35 km/hr.
5. None below, 2 at the limit, 98 above.
6. More drivers coming from the 60 km/hr limit got closer to the 25 km/hr limit than did those from the 100 km/hr zone. Only 16 drivers that came from the 60 km/hr zone traveled at 45 km/hr or more compared to 34 drivers that came from the 100 km/hr zone.

Interaction D

1. A frequency histogram displaying the distribution of weights at which 50 pieces of a cheap brand of fishing line broke.

2. Approximately symmetrical. Most pieces broke at around 1kg, while approximately equal, but decreasing numbers, broke at lesser and greater weights, as low as 0.6kg, and as high as 1.5 kg.

Interaction E

1. We know that around half the pieces broke at approximately 0.8 kg less than the advertised breaking strain.
2. The spread of the distribution.
3. Numerous. eg. That the sample of line was representative of the population of line. That the method used to measure the breaking strain was appropriate or similar to that used by the company.

Interaction F

2.

Weight (kg)	0.6 -	0.8 -	1 -	1.2 -	1.4 -	1.6 -	1.8 -	2 -	2.2 -	2.4 -
frequency f	1	2	1	2	5	4	21	9	4	1

3. Skewed left
4. mean = 1.795 kg, median = 1.875 kg
5. The pieces of mid-range line have more often than not broken at a higher weight than the pieces of cheaper line. A median of 1.05 kg compared to 1.875 kg starts to help us argue that the mid-range price line has a higher breaking strain.
6. The mid-range line's breaking strain is not different from the advertised value, based on the fact that about half the pieces broke above the advertised breaking strain and about half broke below.
7. Information about the spread of the distribution.

Interaction G

1

	60km/hr to 25 km/hr sample	100 km/hr to 25 km/hr sample
shape	skewed right	largely uniform
mean (km/hr)	34.3	40.6
median (km/hr)	32	39

2. The small number of relatively high speeds, illustrated by the skewness, in the first sample cause the mean value to be approximately 2 km/hr higher than the mean. Even though the second sample distribution is largely uniform a similar outcome has occurred. The greater the skewness the greater the difference between the mean and median in the direction of the skewness.

Interaction H

Neither the researcher nor those participating in the research know who is taking the drug and who is taking the placebo.

Interaction I

group	min	Q1	median	Q3	max	IQR
all before	5.00	6.40	7.25	7.90	8.90	1.50
drug Group	3.20	4.60	4.70	5.05	6.20	0.45
placebo group	5.70	6.35	7.20	7.90	8.80	1.55

Interaction J

1. Yes. The cholesterol level distribution for the drug group had median level of 4.7, over 2 units less than the all before group. The interquartile range of 0.45 (4.60 to 5.05), approximately 1 unit less than the all before group, shows the central 50% of values varied less than the all before group and were considerable lower. This, coupled with the fact that the placebo group had a distribution very similar to the all before group supports the conjecture that the drug has been effective in lowering the cholesterol level.
2. The central 50% of speeds for both samples had very similar ranges, the 60 km/hr group having IQR of 13.5 km/hr compared to 12 km/hr of the 100 km/hr group. However, the central 50% of speeds of each group are quite different in value. The 60 km/hr group's values ranging between 26.5 km/hr and 40 km/hr compared to 35 km/hr and 47 km/hr for the 100 km/hr group. The boxplots illustrate this well.
3. The range of the central 50% of each brand's breaking strains is identical (IQR = 0.35 kg). The central 50% of breaking strains of the cheap line were, however, much lower. They ranged from 0.85kg to 1.2 kg compared to 1.65 kg to 2 kg for the mid-value line.
4. This option gives a 'mean boxplot'. The five values illustrated are minimum score, mean minus one standard deviation, mean, mean plus one standard deviation and maximum score.

Interaction K

1. MPA
2. MPA. It's quite likely that there will be two numbers in succession the same at least once.
3. MPA. After 20 throws, there should be a spread of numbers thrown.
4. MPA. You will need to record the results for the whole class to check this. It is likely that each number will be thrown close to one sixth of the total number of tosses, since the six possibilities are all equally likely.

Interaction L

1. MPA.
2. MPA; It's unlikely that there will be patterns in the numbers produced. There will probably be a fairly even spread over the range from 0 to 1.
3. MPA; results are likely to be similar to those obtained with the tables of values.
4. MPA. One way to check is to draw the graph of $y = 0.5$ (using the Y2 function) and counting how many points are on each side of the line.

Interaction M

1. MPA. With only five tosses, it is not unusual for the relative frequency to be 100% (all scores are 1's) or 0% (no scores are 1's).
2. MPA; In fact, you are more likely to get two or four heads than you are to get exactly three heads.
3. Use `Int(Ran# + 1/6)`. MPA, but it's unusual to take longer than 20 attempts.
4. MPA; In the whole class, it's quite likely that someone will get a failure on the first or second attempt.

Interaction N

1. MPA. (It can't be exactly 0.5, since that would represent 63.5 successes, clearly impossible.)
2. MPA
3. MPA, but it is very unlikely that it will drop below 85% for 127 attempts.
4. When a probability of zero is used there will be no successes (impossibility); when a probability of one is used, there will be no failures (certainty). The relative frequency will equal the probability in each case.

Interaction O

1. MPA, e.g. different usage of CD player, differences in manufacture, temperature effects, etc.
2. mean = (41 + 39 + 37 + 42 + 37)÷ 5 = 39.2 hours. (Don't use a calculator to do this!)

Interaction P

1. 31 hours
2. Each face of the die has the same chance of being tossed. So each table has the same chance of being selected. In each table, all 36 scores have the same chance of being chosen.
3. MPA. (Don't use a calculator!) Your mean will probably be between 37 and 42 hours.
4. MPA; They will probably be close, but not too close.
5. MPA. One good idea is to take the mean of all your sample means as an estimate.

Interaction Q

1. MPA. The mean of the sample means is likely to be close to the population mean.
2. MPA. Draw a histogram and find the mean of the sample mean scores.
3. MPA. The sample means should be rather close to each other. They will probably lie between 36 and 38, making them close to population mean.

Interaction R

1. MPA
2. MPA; the sample of size 20 will be more reliable. The means of successive samples are likely to be closer to each other than were the means of the samples of size 5.
3. MPA; test your prediction by comparing it with 36.96. Your sample mean will probably be close to this.

Interaction S

1. MPA
2. Whole numbers between 1 and 8 are generated, since the range for `8xRan#` is from 0 to 8.
3. Use the command `Int(7xRan# + 1)`
4. `Int(6Ran# + 1) + Int(6Ran# + 1)`

Interaction T

1. MPA; About two to four sixes is likely.
2. The results will be different, since they are randomly

generated.

3. Different, since each set is randomly generated.
4. It takes about twice as long to generate twice as many rolls, since each roll takes about the same time.

Interaction U

1. The first few should match exactly.
2. MPA. Notice that the two lists are different.

Interaction V

1. MPA. Both histograms are likely to have peaks near the middle, with modes of 6, 7 or 8.
2. MPA. Similar (but not identical) results to those before are expected.
3. MPA, e.g. the middle scores (e.g., 5 to 9) are more likely than the end scores (e.g., 2, 3, 11, 12), but sometimes this is not the case.

Interaction W

1. MPA. The means should be close to 7 in each case; minimum of 2 and maximum of 12 are likely. Median will usually be 7 and the mode will usually be 7.
2. MPA. Since the mean score is about 7, it will usually take about $40 \div 7 \approx 6$ turns.
3. MPA
4. MPA; the larger sample will usually mean the results are more predictable and so the histogram will be like the one in the text.