Practical 2: Using Minitab

Instructions

1. Read through the instructions below for "Accessing Minitab".

2. Work through all of the exercises on this handout. If you need any help then just ask.

3. This practical is not assessed — however, working through all the exercises here should equip you with the skills needed to use Minitab in **Assignment 1**.

Accessing Minitab

Minitab is an easy-to-use statistical package which can carry out a wide variety of statistical tasks. It can be found on most university computers and can be accessed as follows.

1. Login to a university computer (using your usual university username and password).

2. Minitab is started by clicking on the Windows Start icon, then selecting:
   `All Programs > Minitab > Minitab 17 Statistical Software`

You should now have a spreadsheet ("data window") ready to input data. In Minitab, there are two main windows; the **Session** window and the **Worksheet** window. The Worksheet allows you to view and edit the data columns of the current worksheet. It is normally empty on startup, so the first step is to load the data in. Always enter data in the white boxes — the light grey boxes are for column titles. Use the arrow keys to move around the worksheet.

You can copy-and-paste graphs by right-clicking on the graph in Minitab, selecting `Copy Graph`, *and then, for example, selecting* `Paste` *in Microsoft Word.*

Exercises

1. The observations below are the recorded time it takes to get through to an operator at a telephone call centre (in seconds).

54	56	50	67	55	38	49	45	39	50
45	51	47	53	29	42	44	61	51	50
30	39	65	54	44	54	72	65	58	62

   These data were presented to you in Chapter 2, and you should have drawn a stem-and-leaf plot to represent them on page 7 in the lecture notes. Let us now use Minitab to produce this plot.

   Enter the data in the table above in column `C1` of the Minitab worksheet. To do this, point the cursor at the first (white) cell in column `C1` (corresponding to row 1) and click. Then type in the number 54, move to the cell underneath using the down arrow key, type in the number 56, and so on, until all 30 observations have been entered. Once you have done this, you should give the column a title. To do this, point and click the cursor over the grey box at the top of the column, and type in a suitable title (e.g. "time in seconds").

From the taskbar, click on `Graph > Stem-and-Leaf`, and in the `Graph variables` box type `C1` (because this is the column where your data are). Click `OK`.

The stem and leaf plot should appear in the Session window — check to make sure.

Look at the stem and leaf plot. What comments could you make on the distribution of the call centre data?

2. Also presented to you in Chapter 2 were the data on students' modes of transport to University; these are shown again below.

Mode	Frequency
Car	10
Walk	7
Bike	4
Bus	4
Metro	4
Train	1

Open a fresh Minitab worksheet by clicking on `File > New > Minitab Worksheet > OK`. Enter the data in the table above into columns `C1` and `C2` of the new worksheet, i.e. enter the modes of transport into column `C1` and the corresponding frequencies into column `C2`. Don't forget to give each column a title in the grey box (i.e. "mode" and "frequency"). Notice that column `C1` is now called `C1-T`; this is because Minitab has recognised that this is a **T**ext column.

Click on `Graph > Bar Chart`. Because you have entered the data into Minitab in a tabular form (i.e. the frequency table, and not the raw data, have been entered), in the `Bars represent` box click the down arrow and select `Values from a table`, and click `OK`. In the `Graph variables` box type `C2`, because column `C2` actually contains your data. In the `Categorical variable` box type `C1`, because this column contains the categories. Click `OK` and a bar chart of these data should appear.

Double click on the title of the bar chart. When the `Edit Title` box appears, change the title of this chart to something more appropriate.

Now double click on one of the bars in the chart. In `Attributes` click on `Custom` under `Fill Pattern` and click on the down arrow next to `Background color`. Choose a colour you like and then click `OK`. This will change the colour of the bars in the chart. In the same way, you can change the colour of individual bars.

Why are there gaps between the bars in your bar chart?

3. Open a web browser (e.g. Internet Explorer) and go to the MAS1403 webpage, either directly from the following link

`http://www.mas.ncl.ac.uk/~nlf8/teaching/mas1403`

or via Blackboard.

Click on the tab for "Practicals", and under the heading for "Practical session 2" you will see links to two datasets. Click on the link for the **Farmer's Market Data**; a file will open which shows the amounts, in £, spent by 200 customers at a farmer's market stall in the Lake District. Select the data (by highlighting it – click the left mouse button before the word `Spend` and hold it down as you drag the mouse to the bottom of the page), then right-click the mouse and select `Copy`.

Now go back to Minitab and open a fresh worksheet in exactly the same way you did in Exercise 2. Point and click the cursor on `C1`, which should highlight the entire column. Click on `Edit > Paste Cells`, then click on `Paste as a single column` and then on `OK`; doing so should now enter the Farmer's Market Data into the first column of your new Minitab worksheet.

(a) Produce a histogram of the Farmer's Market Data by clicking on `Graph > Histogram > Simple > OK`. Type `C1` in the `Graph variables` box and click on `OK`. Minitab will display a histogram for the Farmer's Market Data.

(b) Change the colour of the bars in your histogram. Change the title to something more appropriate, and change the label of the x-axis to "Amount spent in pounds".

(c) Now double click anywhere on the histogram, making sure the entire graph is selected. When the `Edit Bars` box opens, click on `Binning`, and in the `Interval Definition` section change the `Number of intervals` to 80 and click `OK`.

(d) Repeat the steps in part (c), but this time change the `Number of intervals` to 5.

(e) Repeat the steps in part (a) to produce the original histogram again. Now double click on any of the numbers on the frequency axis; an `Edit Scale` box should open. Click on `Type`, select `Percent`, and click `OK`. What is the name of the graph now produced?

What is wrong with the histograms produced in parts (c) and (d)? Why might we prefer to use the type of graph produced in part (e)?

4. Produce a frequency polygon of the Farmer's Market Data by clicking on `Graph > Histogram > Simple > OK`. Type `C1` in the `Graph variables` box and click on `Data View`. In the `Histogram: Data View` window, click on `Bars` to deselect it, then click on the `Smoother` panel, click on `Lowess` and then enter 0 in the `Degree of smoothing` box. Then click on `OK` and then on `OK` again. Minitab will display a frequency polygon for the Farmer's Market Data. (Change the title of the graph to something more appropriate in the usual way.)

   If a cumulative relative frequency polygon is required then follow the steps above, but before clicking `OK` for the second time, click on the `Scale` button, then click on `Y-Scale Type`, select `Percent` and `Accumulate values across bins` then click on `OK` and then on `OK` again.

5. Go back to the module webpage and follow the steps in Exercise 3 to load the **Jeans Sales Data** into Minitab. These data are the estimated monthly sales of pairs of Levi's jeans, in thousands, over a six year period in the U.K. (Jan 2008 to Dec 2013 inclusive).

   Produce a time series plot (covered on page 13 of Chapter 2) of these data by clicking on `Graph > Time Series Plot > Simple > OK`. Enter `C1` in the `Series` box.

   Now click on `Time/Scale`, select `Calendar` and from the drop-down menu select `Month Year`. In the `Start Values` boxes enter 1 for `Month` and 2008 for `Year` (i.e. January 2008), and click `OK`, and then on `OK` again.

   Minitab will produce a time series plot for the Jeans Sales Data.

   Look at the time series plot. Can you see any trends? Are there any outliers?

6. Open a fresh Minitab worksheet and re-load the Farmer's market data you used in Exercise 3. We will now see how to use Minitab to produce numerical summaries of data in line with the work we covered in Chapter 3 of the lecture notes.

   Click on `Stat > Basic Statistics > Display Descriptive Statistics`, enter `C1` in the `Variables` box and click `OK`. Minitab will give lots of numerical summaries in the Session window — we're not interested in all of them.

   Use the Minitab output to help complete the table on the next page:

Mean ($\bar{x}$)	
Median	
Range	
Standard deviation (s)	
Inter-quartile range ($Q3 - Q1$)	

   Now use Minitab to produce a box plot by clicking on `Graph > Boxplot`, then under `One Y`, select `Simple > OK`. Enter `C1` in the `Graph variables` box and click `OK`.

   Which is more appropriate as a measure of location for these data, the mean or the median? Why?

7. The work we will do in this question relates directly to material from the lecture earlier this week - it might help if you have your notes open as you proceed!

   Open a fresh Minitab worksheet. We are going to use Minitab to simulate a die-rolling experiment. Suppose you roll a fair, six-sided die **three times**, and you are interested in the number of sixes you observe. In particular, you want to find the probability that you will get:

   (i) No sixes at all;

   (ii) Exactly one six;

   (iii) Exactly two sixes;

   (iv) Exactly three sixes – in other words, all sixes.

   The answer to part (i) should be fairly easy to work out: we can safely assume independence between the outcomes of each roll of the die, and so we want

   $$P(\text{six and six and six}) = P(\text{six}) \times P(\text{six}) \times P(\text{six}).$$

   Now since we have a fair die, with six sides, we know that

   $$P(\text{six}) = \frac{1}{6},$$

   and so

   $$P(\text{six and six and six}) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216} = \mathbf{0.0046} = \mathbf{0.46\%}.$$

The answers to some of the other parts, however, are not quite so easy to work out. For example, the probability of getting exactly one six could mean

$$P(\text{six and six and (not six)}) \quad \text{or} \quad P(\text{six and (not six) and six}) \quad \text{or} \quad P((\text{not six) and six and six}).$$

This is a classic example of the **Binomial distribution**, which we saw in this week's lecture. For now, we will use Minitab to repeatedly simulate our three rolls of the die and observe what happens.

(a) In Minitab, click `Calc > Random Data > Integer`.

In the `Number of rows of data to generate` box, enter **1000**, and in the `Store in column(s)` box, enter **C1–C3**.

In the `Minimum` box enter **1** and in the `Maximum` box enter **6**. Before hitting `OK`, can you guess what Minitab is going to do? Think about this carefully, and perhaps talk to your friends. Now hit `OK`!

(b) In your worksheet, you should have 1000 rows of data, across three columns. Each row represents a single experiment where you roll the die three times; the numbers in the columns represent the outcomes of each of the three rolls of the die. Minitab did this by randomly generating integer values between 1 and 6 across columns **C1–C3** and for **1000** rows. Look carefully through what you did in part (a), all should now make sense!

(c) The next thing to do is "code" the results according to the following rule:

If you rolled a six, that is a 'success' $\longrightarrow$ Replace the outcome with a **1**

If you did *not* roll a six, that is a 'failure' $\longrightarrow$ Replace the outcome with **0**

In Minitab, click `Data > Code > Numeric to Numeric`. In the box which says `Code data from columns`, type **C1–C3**. In the box which says `Store coded data in columns`, type **C5–C7**.

Then in the grid below, where it says `Original values`, type **1:5**, with the corresponding `New` value entered as **0**. This will code all rolls of the die not resulting in a six as the value 0. Similarly, in the next row of the `Original values` column type **6**, with the corresponding `New` value entered as **1**.
Compare columns `C1–C3` with columns `C5–C7`. ***Can you see how the coding has worked?***

(d) Now, for each of your 1000 experiments (each row in the worksheet), you want to find out how many sixes you obtained. Select `Calc > Row Statistics`, then select `Sum`. In the `Input variables` box enter **C5–C7** and in the box which says `Store result in` enter **C9**. Hit `OK`.

Now look at your updated worksheet, and in particular the numbers in column 9. ***Can you tell what Minitab has done?***

(e) Now select `Stat > Tables > Tally individual variables`. In the `Variables` box enter **C9**; select `Percents` and then hit `OK`. Look at the output in Minitab's Session window (the white window above the worksheet). What you should see is an estimate of the **probability distribution** for X : Number of sixes obtained from three successive rolls of the die.

Using this output:

(i) Compare the percentage of 3's to the probability we worked out by hand at the start of this question. What do you notice?

    (ii) If you were to roll a die three times for real, what is the most likely number of sixes you would observe – 0, 1, 2 or 3?

    (iii) What is the probability of obtaining *at least two* sixes?

    (iv) Make a note of the results shown in the Session window (just write them down on a piece of paper), and then run throuh parts (a)–(e) again. **What do you notice?**

(f) Minitab has some very useful built-in functions for calculating probabilities from distributions. In column 11 of your current Minitab worksheet, enter the values **0**, **1**, **2** and **3**. These are the possible outcomes for the number of sixes we obtain when we roll the die three times. Label the column "number of sixes".

Now select `Calc > Probability Distributions > Binomial`; select `Probability` (instead of `Cumulative Probability`); enter **3** as the `Number of trials` (as we are rolling the die three times); enter **0.16667** as the `Event Probability` (the decimal version of 1/6, our success probability); enter **C11** as the `Input column` and enter **C12** in the `Optional storage` box. Hit `OK`.

What has Minitab done here? Compare the results in columns 11–12 with those obtained earlier, an those obtained in the lecture notes from Monday's lecture.

Once you have finished working through the exercises, follow the instructions below for saving your worksheets and exiting Minitab.

Saving and retrieving worksheets and projects

To save your complete workspace, including the session window, all open worksheets, and any plots, select `File > Save Project As`. Make sure that your current drive is `Documents (H:)` in the `Save in:` field, and give an appropriate name for the file before clicking on `Save`. Note that on the university Windows clusters, drive `H:` is synonymous with `My Documents`. This project can be reloaded at a later stage by selecting `File > Open Project` or by clicking on the small yellow "open file" icon on the Minitab toolbar.

Exiting and logging out

When you are finished working with Minitab, it is important that you exit the program and then log out of the Windows network properly; this will ensure that all of your work is saved properly.

To exit Minitab, select `File > Exit`. You will have an opportunity to save your project if you haven't already done so (see the instructions above).

Once you have exited Minitab, you can log out of the cluster by selecting the Windows Start icon and then selecting `Log off`. Don't do this until you have finished all the exercises!