ACC1012/53: Professional Skills for Accounting and Finance

Practical session: Introduction to Minitab

What you need to do

The aim of practical session is to introduce you to the statistical analysis software Minitab. Today's class is an elementary introduction to Minitab 17, the version of this software that is installed on computers in all of the University's PC clusters. A more comprehensive instruction manual is available to download from the ACC1012/53 webpage, in the Minitab section.

To load Minitab follow the steps below; then work through the questions overleaf, which all relate to material in Chapters 3–5 of the lecture notes. You can access Minitab via RAS, or you can even buy your own copy from ISS (about £10). Future computer practicals will make use of Minitab's ability to perform more sophisticated analyses.

1. Login to a PC (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
3. Via RAS, you will find Minitab in the Statistical Software folder.

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.

Some of this work is assessed, and will go towards your overall mark for the module. The starred questions from this practical, as well as those from the next practical session, should be submitted together as a single piece of work by **4pm, Friday 20th April 2018**. As always, your work should have a cover sheet attached and should be posted through the homework submission letterbox on the 3rd floor of the Herschel Building.

Warning: Don't just work through the assessed questions! Make sure you work through *all* of the questions! We can give you as much help as you need with the non–assessed questions, and you might struggle to answer the starred questions without having worked through the other questions first.

See the end of this handout for information on saving Minitab worksheets. You can copy-and-paste any of the graphs you produce in the following questions by right-clicking on the graph in Minitab, selecting Copy Graph, and then, for example, selecting Paste in a word-processing application such as Microsoft Word.

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 3 (Example 3.3) and you should have drawn a stem-and-leaf plot to represent them on page 67 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 cell in column `C1` 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` $\rightarrow$ `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?

How does this plot differ from that drawn in the lecture? Can you work out what Minitab does differently?

2. Also presented to you in Chapter 3 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` $\rightarrow$ `New` $\rightarrow$ `Minitab Worksheet` $\rightarrow$ `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.

We will now produce a bar chart and a pie chart for these data (please turn over).

(a) First of all, the bar chart.

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 on Chart Options and then select Decreasing Y, which will put your bars in order of frequency (from tallest to shortest). Now click OK and OK again, 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?

(b) Now for a pie chart.

Click on Graph → Pie Chart. Now follow similar steps to those given in part (a) for the bar chart to produce your pie chart.

3. Open Internet Explorer and go to the ACC1012/1053 webpage, either directly from the following link

http://www.mas.ncl.ac.uk/∼nlf8/teaching/acc1012/

or via Blackboard.

Click on the Minitab link in the taskbar, and scroll down to the section on "Computer session 1". Here, you should see links to some 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. In Internet Explorer, right-click anywhere in the main window, then select Select all, then right-click again and select Copy.

Now go back to Minitab and open a fresh worksheet in exactly the same way you did in Question 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 on any one of the bars. 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. Go back to the module webpage and follow the steps in Question 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 2006 to Dec 2011 inclusive).

Produce a time series plot (covered on page 78 of Chapter 3) of these data by clicking on `Graph` $\rightarrow$ `Time Series Plot` $\rightarrow$ `Simple` $\rightarrow$ `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 2006 for `Year` (i.e. January 2006), 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?

5. Open a fresh `Minitab` worksheet and re-load the Farmer's market data you used in Question 3. We will now see how to use `Minitab` to produce numerical summaries of data in line with the work we covered in Section 3.5 of the lecture notes.

Click on `Stat` $\rightarrow$ `Basic Statistics` $\rightarrow$ `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 following table:

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

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

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

6. Recall question 9 from Section 4.8 of Chapter 4:

> Between 1896 when the Dow Jones Index was created and 2009, the index rose in 64% of the years (data taken from M. Hulbert, "What the Past Can't Tell Investors", *New York Times*, January 3rd, 2010). Based on this information, and assuming a binomial distribution, estimate the probability that the stock market will rise (a) In exactly four of the next five years; (b) In none of the next five years.

(a) In Chapter 4, we looked at how to use the formula for the binomial distribution to answer such questions. For example, for part (a):

$$P(X = 4) = {}^5\text{C}_4 0.64^4 0.36^1 = 0.3020,$$

where X : the number of years the stock market rises. In `Minitab`:

(i) Click on `Calc` → `Probability Distributions` → `Binomial`, and then select `Probability` (as opposed to of `Cumulative probability` or `Inverse cumulative probability`).

(ii) Now enter the `Number of trials` as 5 and the `Event probability` as 0.64.

(iii) Finally, select the option `Input Constant` and enter 4 – because in this example, we want to find the probability that the stock market rises in exactly the next four years, i.e. $P(X = 4)$.

When you hit `OK`, you should get the following output:

Probability Density Function

```
Binomial with n = 5 and p = 0.64

 x     P(X=x)
 4  0.301990
```

Notice that `Minitab` gives the same answer as we got by hand, correct to 3 decimal places.

(b) Follow the example above to obtain the probability that the stock market will rise in *none* of the next five years – confirm that this gives the same as the answer you obtained from question 9 in Section 4.8 of the lecture notes.

(c) We can use `Minitab` to obtain the *full probability distribution* of our random variable X. In terms of the stock market example, note that X can take any value in $0, 1, \ldots, 5$. We *could* work out the probability of each outcome individually, either by hand or as we did for $X = 4$ and $X = 0$ in parts (a) and (b) above. However, `Minitab` can be used to work out all of these probabilities together.

(i) In column C1 of your `Minitab` worksheet, enter the full list of possible outcomes for X, that is:

	C1	C2	···
1	0		
2	1		
3	2		
4	3		
5	4		
6	5		
7			
⋮			

(ii) Now select `Calc` → `Probability Distributions` → `Binomial`, and make sure `Probability` is selected (rather than `Cumulative probability` or `Inverse cumulative probability`).

(iii) As before, make sure you enter the `Number of trials` and `Event probability`.

(iv) Now select the `Input column` option, and enter `C1`.

(v) Click `OK`.

Now you should see that `Minitab` has computed the full probability distribution for X, that is, the binomial probability for each of the possible outcomes for X.

(d) Suppose we want to know the probability that the stock market will rise in fewer than 3 of the next five years, i.e. $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$. Of course, you could just add together the corresponding probabilities from your answer to part (c) above; however, `Minitab` can also be used to obtain such *cumulative probabilities*.

(i) As before, select `Calc` → `Probability Distributions` → `Binomial`.

(ii) Select `Cumulative probability`.

(iii) Make sure your `Number of trials` and `Event probability` have been entered, as before.

(iv) Enter `C1` as the `Input column`.

(v) Click `OK`.

Now you should see:

Cumulative Distribution Function

```
Binomial with n = 5 and p = 0.64

 x  P(X<=x)
 0  0.00605
 1  0.05979
 2  0.25090
 3  0.59064
 4  0.89263
 5  1.00000
```

Using this output, what is $P(X < 3)$?

(e) Use your answer to part (d) to find $P(2 \leq X \leq 4)$.

[Please turn over for more questions]

7. More recently, we have looked at the Normal distribution as a probability model for continuous data. Recall the example about the *standard* Normal distribution from last week's lecture (pages 126-127 of the lecture notes). Here, we have $Z \sim N(0, 1)$, and we used tables to find:

1. $P(Z < -1.46)$
2. $P(Z < -0.01)$
3. $P(Z < 0.01)$
4. $P(Z > 1.5)$
5. $P(-1.2 < Z < 1.5)$

We will now use `Minitab` to work out thes probabilities.

(a) First of all, $P(Z < -1.46)$. Click on `Calc` → `Probability Distributions` → `Normal`; select `Cumulative probability`; enter the `Mean` as 0 and `Standard deviation` as 1; and then enter the `Input constant` as `-1.46`. Clicking `OK` should give $P(Z < -1.46) = 0.0721450$ – the same as we got in the lecture, to 4 decimal places.

(b) Repeat what you did in part (a) to find $P(Z < -0.01)$, and check this with your answer in the lecture notes (page 123).

(c) Repeat what you di in parts (a) and (b) to find $P(Z < 0.01)$, and check this with your answer in the lecture notes.

(d) Use `Minitab` to find $P(Z > 1.5)$. What will you have to remember to do here? Check your answer with that in the lecture notes.

(e) Use `Minitab` to find $P(-1.2 < Z < 1.5)$. Again, check your answer with that in the lecture notes.

8.* This is an assessed question.

For this question, you must analyse your own unique dataset. *Your mark for this question will be zero if you use another student's dataset!*

To access your own personal dataset for this question, go to the course website – via Blackboard, or directly at:

`www.mas.ncl.ac.uk/~nlf8/teaching/acc1012/`

Click on the Minitab link in the taskbar, and then scroll down to the section for "Computer session 1". Now click on the link that says "Personal datasets". A new page should open, and you should then click on your name and print off your personalised datasets for question 8. You must use `Minitab` to answer this question – any graphs or output from `Minitab` should be copied–and–pasted into a Microsoft Word document (see the end of this handout). You may find the supplementary `Minitab` helpsheet, on the course webpage, helpful for producing plots that we have not covered in this worksheet.

[Make sure you write down your dataset number in your solutions – this should appear at the top of the supplementary sheet]

(a) Produce appropriate graphical summaries for your data.

(b) Produce appropriate numerical summaries for your data.

(c) Write a short paragraph explaining your findings.

9.* This is an assessed question.

Recall question 12 from Section 4.8 of the lecture notes:

> In 2012, the airline *Cathay Pacific* "mishandled" 3.89 bags per 1,000 passengers. What is the probability that, in the next 1,000 passengers, *Cathay Pacific* will have (a) no mishandled bags; (b) at least one mishandled bag; (c) exactly two mishandled bags?

You should have realised that we now have a Poisson distribution, that is, $Y \sim Po(3.89)$, where Y : Number of mishandled bags (per 1,000 passengers). By hand, we can work out the answer to part (a) by using the Poisson formula, i.e.:

$$P(Y = 0) = \frac{e^{-3.89}3.89^0}{0!} = 0.0204.$$

(a) Follow what you did in question 6(a) to find the probability that, in the next 1,000 passengers, this airline will have no mishandled bags, that is $P(Y = 0)$. *Include in your solutions any relevant output from* `Minitab`.

(b) Following what you did in part (a), find the probability that, in the next 1,000 passengers, the airline will have exactly 2 mishandled bags. *Include in your solutions any relevant output from* `Minitab`.

(c) Find the probability that, in the next 1,000 passengers, *Cathay Pacific* will mishandle fewer than six mishandled bags. *Include in your solutions any relevant output from* `Minitab`.

Once you have finished working through the questions, follow the instructions overleaf for saving your worksheets and exiting `Minitab`.

Saving and retrieving worksheets and projects

When you have been using Minitab, you will often want to save the contents of your Worksheet for future use. To save a Worksheet, first click on it in order to make it active, then select File → Save Current Worksheet 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.

Note that saving a Worksheet only saves the Worksheet contents. It does not save any plots you have produced, or the contents of the session window. To save your complete workspace, including the session window, all open worksheets, and any plots, select File → Save Project As and select an appropriate folder and file name. This 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. Projects are often more convenient than worksheets for a "project" you are working on. However, they are less useful for long term data storage, as the project files tend to be very large.

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 and that your files are not corrupted.

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 the end of your practical session!