QUEEN MARY, UNIVERSITY OF LONDON

MTH 4106 Introduction to Statistics

Practical 1 10 January 2012

In this practical you will be introduced to the statistical computing package called Minitab. You will use this package throughout this module, and also in some later modules. It is important that you learn how to use it. Some assignment questions will require use of Minitab. Some test and examination questions may require you to interpret output from Minitab.

The work is on the student PC network. I assume that you know how to log onto the PC network and how to print files from it.

1 (Getting Started) Before you log on, you need to choose the

`Standard Windows Environment`

After logging on, choose

`Public >>` `Change Course Environment >>`
`Mathematics Department >>` `Statistics I >>`

The first time that you use Minitab for *MTH 4106 Introduction to Statistics* you should create a new directory (also called *folder*) in your own space and call it `STATS`. First move to your own space, by clicking on the icon for **My Filestore**.

In `File and Folder Tasks`, choose `Make a new folder`

A folder icon will appear in the window. Delete its initial name "New Folder" and give it the new name "STATS". Close the window (click on × in the right top corner of this window).

Remember to store all your files and results from this module in `STATS`.

2 (Starting Minitab) To start Minitab,

`Dept >>` `Minitab-15.1` ▶ `Mtb`

It is good to start your encounter with Minitab through a Minitab Tutorial. For this practical, work through the following tutorial, which is based on one previously supplied with Minitab. Make notes on what you learn from this tutorial. Also, use the empty boxes to write answers to the questions posed during the tutorial.

The story Clones are genetically identical cells descended from the same individual. Researchers have identified a single poplar clone that yields fast-growing, hardy trees. These trees may one day be an alternative energy resource to conventional fuel.

Researchers at The Pennsylvania State University planted trees of Poplar Clone 252 on two different sites: one site was by a creek, with rich, well-drained soil, and the other site was on a ridge, with dry, sandy soil. They took a sample of three-year-old trees and measured their diameter in centimetres, height in metres, and dry weight of the wood in kilograms. These researchers want to see if they can predict how much a tree weighs from its diameter and height measurements.

Congratulations! You have been hired as data analyst for the project, and you will be performing the statistical analysis.

In this session you will learn how to:

- open a worksheet
- enter and edit data
- save data
- compute some basic statistics
- do arithmetic
- plot the data
- edit and add comments to the output
- print and save your results.

Step 1 Start Minitab.

Step 2 Open a Worksheet.

When you start Minitab, you begin with a new, empty project. You can add data to your project in many ways, but the most common way is to open a worksheet. Note that you are only copying the data from the worksheet to the project; any changes that you make to the data added to your project will not affect the original file.

In this session, you will use the file `POPLAR1.MTW`. This file is one of the dozens of worksheets that are supplied with Minitab. Most of these worksheets are in the `Minitab Sample Data` subdirectory or folder.

1. Activate the **Project Manager** by typing `[Ctrl]+[I]` or by clicking the `Project Manager` button on the toolbar.
2. Right-click on the Worksheets folder in the **Project Manager** and choose `Open Worksheet...`.

3. Click on the icon `Look in Minitab Sample Data folder`.
4. Select `Poplar1.MTW`, click `Open`, and then click `OK`.
5. If the **Data window** is not already visible, open it to view the columns in your worksheet: either choose `Window` > `Poplar1.MTW` or press `[Ctrl]+[D]`.

This worksheet contains three variables, labelled `Diameter`, `Height`, and `Weight`. Each variable contains 15 observations—all the data collected so far. Look at the **Data window**. How tall is the 14th tree?

Step 3 Enter Data from the Keyboard.

The worksheet `Poplar1` contained the data collected so far, but you have just received new observations from the field, and there are five new rows to enter.

1. Press arrow keys until you reach the first blank cell in row 16 or, with your mouse, click on the first blank cell in row 16.
2. Make sure the data entry arrow to the left of `C1` points to the right. If it does not, click on it to change its direction. Look at the **Data Window** and make sure that this arrow points to the right.
3. Type the following from left to right across each row:

```
1.52  [Enter]  2.9   [Enter]  .07  [Ctrl]+[Enter]
4.51  [Enter]  5.27  [Enter]  .79  [Ctrl]+[Enter]
1.18  [Enter]  2.2   [Enter]  .03  [Ctrl]+[Enter]
3.17  [Enter]  4.93  [Enter]  .44  [Ctrl]+[Enter]
3.33  [Enter]  4.89  [Enter]  .52  [Ctrl]+[Enter]
```

   Tip If you make a mistake: click on or move to a cell (the contents will be automatically selected), type the correct value, and press `[Enter]`.

Step 4 Enter Patterned Data

You now want to create a new variable that will indicate whether an observation was taken from the site with rich, well-drained soil (Site 1), or from the site with dry, sandy soil (Site 2). This new variable, called `Site`, will contain ten 1s followed by ten 2s.

You can always type data in the **Data window**, but if your data follow a pattern, you may find it easier to use the `Autofill` feature to create patterned data.

1. Go to column `C4` and select the `Label` cell, directly under the column number.
2. Type `Site` and press `[Enter]`.
3. In row 1, type `1`.
4. Place your mouse over the square handle in the lower right corner of the selected cell (where you have just typed '1'). The crosshair cursor should change from white to solid black, which indicates that you are in `Autofill` mode.
5. *Left*-click and drag the cursor down to row 10. `Autofill` will automatically fill the selected cells with 1s.
6. In row 11, type `2`.
7. Click with the `Autofill` cursor and drag down to row 20.

Step 5 Save Your Project.

It is a good idea to save your work frequently. Now is probably a good time to save, since you have just entered new data.

1. In the **Project Manager**, right-click on the top-most folder (which is labelled "Untitled") and choose `Save Project As ....`
   Make sure that you are in your folder ***`STATS`*** *before saving the project. You may need to move there using the icon* ***My Filestore*** *or* ***My Computer****.*
2. In `File name`, enter `Poplar1` for the name of your project. If you omit the extension `.MPJ`, Minitab will automatically add it once you save the document.
3. Click `Save`.
4. If you see a message box asking if you want to replace an existing file, click `Yes`.

Once you have saved your project, the project folder in the **Project Manager** will be labelled `Poplar1.MPJ`.

Step 6 Compute Descriptive Statistics.

Minitab offers a wide array of basic statistics to help you analyse your data, such as descriptive statistics, z-tests, and correlations. You decide to produce summary tables and boxplots describing the variables Diameter, Height, and Weight for the trees at each site.

1. From the menu bar at the top of the screen, choose

   `Stat` > `Basic Statistics` ▶ `Display Descriptive Statistics`.

2. In the variable list box, click `Diameter` and drag the mouse so that you highlight `Diameter`, `Height`, and `Weight`. Then click `Select`.

   Note When you select a series of columns, Minitab uses a dash to abbreviate the series. In this example, `Diameter-Weight` means the variables `Diameter`, `Height`, and `Weight`.
3. Click in the `By variables` box, then highlight `Site` and click `Select` again. (Don't click `OK` yet!)

   This tells Minitab to generate separate statistics for `Diameter`, `Height`, and `Weight` for each level of the variable `Site`.
4. Click `Graphs...`.
5. Tick `Boxplot of data` and click `OK` in both dialogue boxes.

Minitab displays text output in the **Session window** and each graph (three, in this case) in its own **Graph window**. Look at the content of these four windows, and write a short summary of what Minitab has put there.

Step 7 Look at the Graphs.

Now you can tile the graphs to view all of them at the same time on your screen.

1. Activate the **Project Manager** by typing `[Ctrl]+[I]` or by clicking the `Project Manager` button on the toolbar.
2. Click on the `Graphs` folder.
3. In the right pane of the **Project Manager**, select the graphs you want. You can select multiple graphs by clicking on the graph titles while holding down the `[Ctrl]` key, or by clicking below or to the side of the graph titles and dragging the cursor across the desired selections.
4. Right-click on the selected graphs and choose `Tile`.

Judging from the boxplots, which site has the larger poplars?

Step 8 Look at the Descriptive Statistics

Use the toolbar or `[Ctrl]+[M]` to obtain the **Session window**. Write down the median values for `Diameter`, `Height`, and `Weight` at each site. Does this confirm your previous conclusion about the site with the larger poplars?

	Site 1	Site 2
Diameter		
Height		
Weight		

Which variable has the largest standard deviation relative to its mean?

What are the minimum and maximum weights at Site 2?

Minimum	Maximum

It appears that some of our poplars are doing very well, while others are barely alive.

Step 9 Perform Arithmetic.

Now on to the task of predicting how much the trees weigh. Based on previous work, the researchers have found that the weight of a tree is closely related to the square of its diameter multiplied by its height. Since you have diameter and height data, you can calculate this new variable by using Minitab's calculator. This uses the formula which you enter and puts the result in the variable which you specify.

1. From the top menu bar, choose `Calc` > `Calculator`.
2. You decide to call the new variable "`D2H`" for diameter squared times height. In `Store result in variable`, type `D2H`.
3. In `Expression`, type `C1**2*C2`. Click `OK`.

   This expression tells Minitab to square the variable `Diameter` (`C1`), then multiply by the variable `Height` (`C2`), and put the result in a new variable called `D2H`.

Tip You could also use the mouse to create the equation: (1) select `Diameter` from the variable list, (2) click the `**`, `2`, and `*` buttons on the calculator, and (3) select `Height` from the variable list.

The **Data window** shows the new variable `D2H` that you just created. If you have lost this window, you can retrieve it from the toolbar or by pressing `[Crtl]+[D]`.

What is the value of the variable `D2H` for the tree in row 3? How was this value calculated? Does the result look approximately correct?

It is useful to do occasional checks like this to make sure that you have not made any mistakes.

Step 10 Save the Project again.

Now save the project changes.

Choose `File` > `Save Project`, or press `[Ctrl]+[S]`.

Step 11 Create a Scatterplot

The researchers have determined that there is a relationship between weight and D2H. You want to see if your poplars' data also exhibit this relationship, so you need to plot `Weight` against `D2H` on a scatterplot.

1. Choose `Graph` > `Scatterplot...`.
2. Choose `With Groups`, then click on `OK`.
3. In `Y` (the vertical axis), enter `Weight`.
4. In `X` (the horizontal axis), enter `D2H`.
5. In `Categorical variables for grouping` enter `Site`, click `OK`.

Looking at the scatter plot, you see a positive linear relationship between `Weight` and `D2H`. That is, as `D2H` increases, so does `Weight`. You also notice an unusual data point—a tree that has a very low weight for a relatively high `D2H` value. For now, you decide to ignore it, but it is something you may want to check on later.

Note In `Graph` > `Scatterplot` you can customize your graphs : for example, you can add a title or change the plotting symbol.

Step 12 Write a Minitab Project Report

It is possible to print separately some or all of the **Session window** and some or all of the graphs that you have produced. (Look at the first tutorial under `Help` > `Tutorials` if you need to do this.) However, it is generally better to put any work that you wish to save into a Minitab **Project Report**.

You may create a report of your work using the `ReportPad` folder in the **Project Manager**.

1. Choose important pictures and put them into your report by right-clicking within the graph and selecting `Append Graph to Report`. (This may be slow—be patient!)
2. Choose important results from the **Session window** and put them into the report by highlighting the relevant lines, right-clicking and then selecting `Append Selected Lines to Report`.
3. If you cannot see your **Project Report**, bring it into the foreground by clicking on `ReportPad` in the **Project Manager**. Write some relevant comments about the results and the pictures in the **Project Report**.
4. Save your **Project Report** in your folder as an `.rtf` file. To do this, right-click on `ReportPad` and `Save Report As ...`

If you need to print a report you can do so by `File` > `Print Report`.

Step 13 Save Your Work.

When you save your project, you save all your work at once: all the data, all the output in the **Session window**, and all the open **Graph windows**. When you reopen the project, all that information will be waiting for you, right where you left it.

1. Choose `File` > `Save Project`.

In fact, if you want to use output or data in another application or another Minitab project, you can save your **Session window** output, data, and graphs as separate files. These separate files are copies of what is currently in your project—the contents of your project are not changed in any way.

Step 14 Exit Minitab.

1. Choose `File` > `Exit`.
2. Minitab may ask if you want to save changes to your project. Since you already saved your project above, there is no need to do it again here. Click `No`.

NOTE: I am not asking you to hand in any part of this work. It is important for future weeks' assignments that you know how to do what has been described in this practical.