MATH2019 – Statistics for Civil and Environmental Engineering

Worksheet 1 – Introduction to MINITAB

The purpose of this worksheet is to get you started using MINITAB. It is fairly self-explanatory, and I recommend you work through it. You might also want to try reproducing some of the summaries and graphical displays from Chapter 1 of the course notes.

What is MINITAB?

MINITAB is a general purpose statistical software package. Its functions range from elementary exploratory data analysis to quite sophisticated statistical modelling. MINITAB is **not** a spreadsheet package, and in comparison with spreadsheet packages its data handling and graphing facilities may seem a bit cumbersome. However its statistical facilities are generally more reliable and far more comprehensive.

Preliminaries

It is assumed that you are a registered user of the Computing Services system, and are familiar with using the Windows machines in the Computing Services clusters. Operations you should be familiar with include logging in, accessing programs and files, file handling, using disks, printing, logging out *etc.* If you are unfamiliar with any of these, then Computing Services provide a wide range of introductory printed material, available in many of the clusters, to help you.

Course Files

A number of data files, which we shall use during the course, are stored on the University central file server. To make these files available on the R: drive of your computer, follow the path All **Programs** \rightarrow **Course Folders** \rightarrow **Access to Course Folders**. The R: drive is opened and the course folders are immediatly visible. You can access the required data files directly from folder math2019 on the R: drive, or copy this folder onto a disk, or into your personal storage space (H: drive) for future use.

Starting MINITAB

Start up MINITAB by following the path All Programs \rightarrow Statistics \rightarrow Minitab. After a short delay a large MINITAB window containing two smaller windows will appear. One is the *Session* window which contains any text output. You can also type commands directly into this window but you first need to choose **Enable commands** from the **Editor** menu, after which the MINITAB prompt (MTB >) will appear. Most commands will be executed using the pull-down menus.

The other window is called the *Worksheet*. This window contains data to be analysed using MINITAB. Notice that when you click on a particular window, then the title of that window is highlighted, and any input from the keyboard is made there. This is referred to as the window being *active*.

MINITAB has an extensive on-line help system. You can access this using the **Help** menu. The help system is particularly useful for looking up commands or functions that you know exist but whose name or whose syntax you have forgotten.

Entering Data in the Worksheet

The following table gives experimental measurements, made by two alternative methods, A and B, on the latent heat of water [Source: Natrella, M. (1963). *Experimental Statistics*. National Bureau of Standards Handbook **91** (Washington, D.C.)]:

Method A	79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97
	80.05 80.03 80.02 80.00 80.02
Method B	80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

The easiest way to enter such data into MINITAB is to enter it directly into the worksheet. To enter data anywhere in the worksheet, simply move to the required position and enter new data or overwrite what is already there. Use the mouse or the arrow keys $(\leftarrow \uparrow \downarrow \rightarrow)$ to move around. MINITAB uses columns (C1, C2, C3, ...) to store data. Enter the measurements for method A into positions 1–13 of C1 and the measurements for method B into positions 1–8 of C2. Note that a number is entered in the worksheet only when you leave the cell in which it has been typed.

In addition to entering data directly, a number of commands are available in the **Edit** and **Data** menus when the data window is active. These are particularly useful for editing a column that contains errors or omissions. You can use these commands to insert or delete individual cells, or whole rows or columns.

Columns in the worksheet can be given names, which are stored below the labels $C1, C2, \ldots$ at the head of the columns (use the mouse to get there). Give the names Method A and Method B to columns C1 and C2. MINITAB uses a * to denote missing observations.

Entering Data from a File

If you have data which is stored in a file on a disk, then you can enter that data into the worksheet without typing in all the numbers by hand. For example, the file wave.dat contains two columns of data collected as part of a study to establish how the choice of mooring method for a wave energy device affected the bending stress in part of the device. For each of 18 simulated sea states, the bending stress (in Newton metres) in the device was measured for each of two mooring methods.

This data may be entered into a worksheet directly by choosing **Open Worksheet** from the **File** menu. A *dialogue box* will appear which will prompt you for a file name. If the data file is a previously saved MINITAB worksheet, then the file will have the .mtw suffix. Enter the name of the file in the space provided and click OK. Data from an ordinary text file (.txt or .dat suffix) may also be entered by choosing **Open Worksheet** from the **File** menu. However you need to click **Options** in the dialogue box, and set **Variable Names** to None and set **Field Definition** to **Free format**, before entering the name of the file. The **Open Worksheet** command allows you to search for the file you want to read, which might be in R: \math2019 or on a floppy disk in the A: drive.

Note that each time you read data from a new file, MINTAB creates a new worksheet (data window). When performing statistical analysis, only the data in the front window is available. You may switch between data windows by using the **Window** menu.

Having successfully entered some data, it is often a good idea to save it to disk in case anything should go wrong during the session and you need to start again. To save the current worksheet you should choose **Save Current Worksheet As** from the **File** menu. A dialogue box will appear which will prompt you for a file name and location. The worksheet file can then be entered in a future MINITAB session using **Open Worksheet** as described above.

A Simple Graphical Display

A simple display for a small set of data in a single column is a *dotplot*, which represents each number by a dot against a measurement scale. To get a dotplot of the latent heat measurements for method A, make sure the relevant worksheet is at the front and choose **Dotplot** from the **Graph** menu, followed by One Y/Simple, and click OK. Enter C1 or Method A under **Graph variables** in the dialogue box and click OK. Do the same for Method B.

If our aim is to compare the two methods of measurement, A and B, the dotplots above are not very useful since they are presented on different scales. To get dotplots of the two columns on the same scale, choose Multiple Y's/Simple and enter both C1 and C2 under Graph variables in the dialogue box.

It appears from the plots that perhaps there is some systematic difference between methods A and B. This could be explored further using a formal statistical test, such as the two-sample t test (to be covered later).

A better quality graphical comparison of the two samples is provided by plotting a boxplot of each sample, on the same axes. Choose **Boxplot** from the **Graph** menu, followed by Multiple Y's/Simple, and click OK.

Enter the names of the two columns in the **Graph variables** box (Graphs 1 and 2 respectively) and click OK. Text and shading can be altered or added by using the **Editor** menu.

Arithmetic and Statistical Summaries

MINITAB has many facilities for arithmetic on columns in the worksheet, including standard statistical summaries. Try the following:

MTB > mean c1 MTB > describe c1 c2

The command DESCRIBE provides several summary statistics for a column, including its mean, median, standard deviation, minimum and maximum. As usual, it is easier to perform this command using the pull down menus.

Choose Basic Statistics \rightarrow Display Descriptive Statistics from the Stat menu. A dialogue box will appear, and you will need to enter the names of the variables for which you require summary statistics under Variables in this box, and then click OK to complete the operation.

Simple arithmetic is done using the LET command, in which the result of calculations on the righthand-side is assigned to a column. For example, to create a column containing the difference in the bending stresses for the two mooring methods type

MTB > let C3=C2-C1

in the session window. Alternatively, calculations may be performed on scalars (single numbers which are stored in MINITAB using K1, K2, \ldots) *e.g.*

MTB > let k1=mean(c2)-mean(c1)
MTB > print k1

The PRINT command may be used to view one or more scalars, as above, or columns, e.g.,

MTB > print c1 c2

Simple arithmetic and statistical operations on data columns are also available using the **Calc** menu.

A Simple Plot

To examine the association between the bending stresses for the two mooring methods, it is useful to produce a simple scatterplot of these two variables. To do this you should choose **Scatterplot** from the **Graph** menu, followed by **Simple**, and click OK. A dialogue box will appear, and you will need to enter the names of the variables you want to plot, as the **Y** and **X** as **Variables**, and then click OK.

A plot will appear in its own window. If you want to produce a hard copy of the plot then you should choose the **Print Graph** option from the **File** menu while the plot window is active. A copy of the plot will be sent to the print queue. You can remove the plot window by clicking on the button at the top right hand corner.

Saving the Session Window

The Session window may be printed at any time, by choosing **Print Session Window** from the **File** menu while the session window is active. However, it is usually best to wait until the MINITAB session is complete, then save this window separately using **Save Session Window As** from the **File** menu. Again, you will be prompted for a file name and a location in which to save the file. The saved Session window file is suitable for viewing on screen, for printing, or for editing with a text processor (*e.g.* Microsoft Word). In particular, you can use a text processor to delete any unwanted lines, such as lines containing errors, and to incorporate extra explanatory text.

Saving your Work and Finishing Off

Before quitting MINITAB, you may want to save your work so that you can carry on where you left off. In §0.6 we described how to save a worksheet. However, it is possible to save the whole of your MINITAB session in a single file: the session window, all the worksheets, and any graphs which you have not removed.

These are saved as a MINITAB project file (.mpj suffix). To save a MINITAB project you should choose **Save Project As** from the **File** menu. Again, you will be prompted for a file name and a location in which to save the file. This will usually be the C: drive of the computer you are using, the A: drive (floppy disk) or the H: drive (your central file storage). Note that any material in the C: drive may be lost once you log out.

A saved worksheet (.mtw) or project (.mpj) file is not suitable for viewing and should only be used within MINITAB. If you want to use an existing MINITAB project file in a new MINITAB session, then you can open the file using **Open Project** from the **File** menu.

You can quit MINITAB by clicking on the button at the top right hand corner of the main MINITAB window. MINITAB will warn you if you have made any changes since you last saved the current project.