*Spatial Data & Analysis (GRG310C)*

**Introduction to the EXCEL package**

**Descriptive Statistics**

Excel is a software program designed for the presentation and analysis of information in a spreadsheet format.

1. **Definitions**

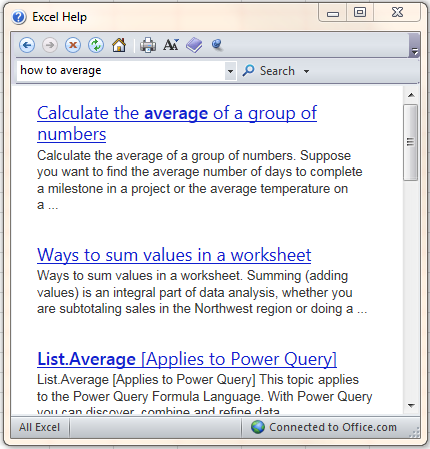
These are specialist words and general terminology which the Excel package uses. Make sure you are familiar with them before trying to use the package.

* Excel – a spreadsheet software package with many functions and facilities that permit the user to undertake statistical and graphical activities
* Spreadsheet – a table or matrix consisting of a series of rows and columns within which data can be recorded and manipulated
* Worksheet – a single page of a spreadsheet, analogous to a single page in a book
* Workbook – a complete spreadsheet used for organizing groups of worksheets. The maximum number of worksheets that a workbook can have is 255

Excel is a flexible way of storing, manipulating, analyzing and displaying data. It is also a useful way to provide data entry for other software packages (such as SPSS which we will be using later in the semester). There are often multiple ways of achieving a particular result in Excel, so as long as you are getting the correct result, then you are doing it right!

\*\*\*Remember to save your work regularly, you don’t want to re-do all your data entry and analysis\*\*\*

1. **Help**

If the Excel package seems intimidating, don’t worry, there is plenty of help at hand.

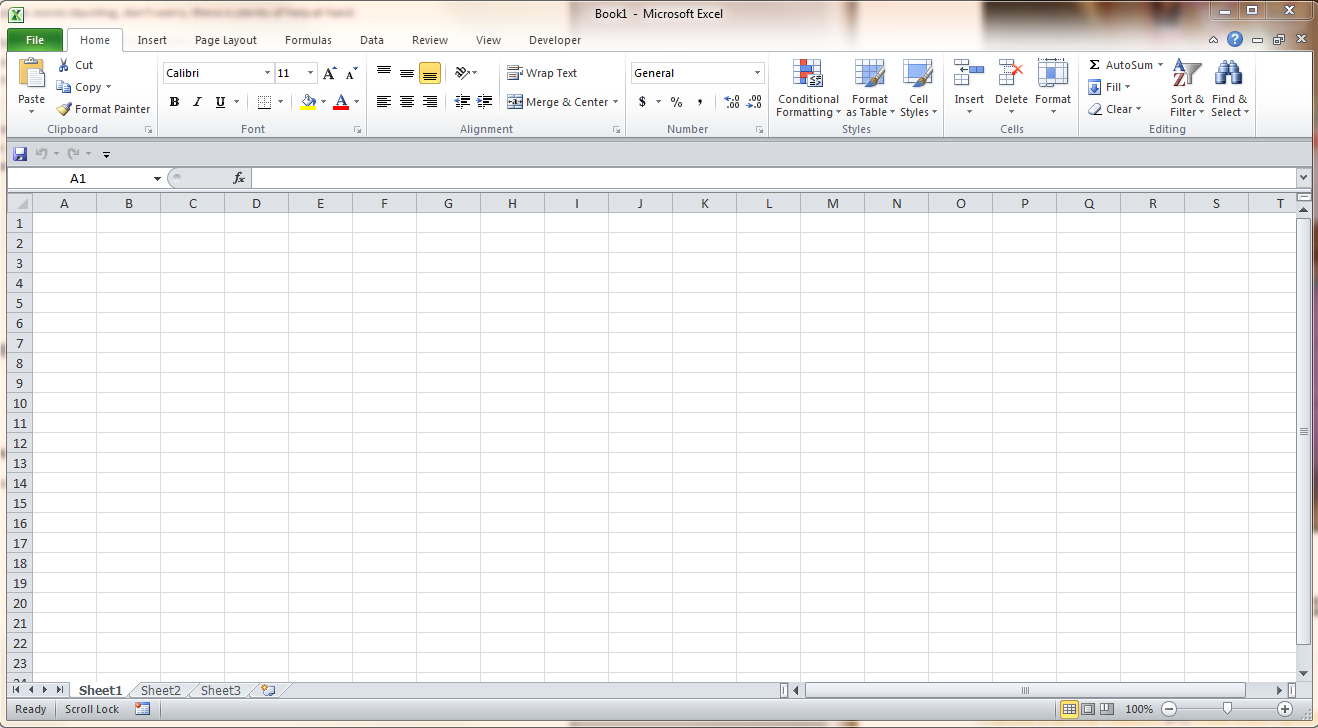
**File > Help** - This option provides help in a variety of formats including where to start, and specific queries. Type your question into the search bar, and you will get a selection of possible options from the web.

Get into the practice of using the help function or at least Googling your issue first. And remember, I provide most of the instructions needed, so make sure you **follow the instructions.**

1. **Getting Started**

Log onto the computer, and click on the Excel icon.

**Title Bar Formula Bar**

**Menu Bar**

**Toolbars**

**Name Box**

**Row Headings**

**Cell Worksheet Scroll Bar Zoom**

**Title Bar –** At the top of the screen, and says the title of the file, i.e. *Book 1 – Microsoft Excel*

**Menu Bar –** The standard Windows display for toolbars

**Toolbars –** Windows tools that are incorporated into Excel

**Formula Bar –** This consists of two parts. The Name Box, a white box which are identified if there are any active cells. There is also a Formula Bar which indicates the formula or value that you have selected to be applied to the data identified in the Name Box.

**Row Headings –** identifies each of the rows by a number (1;2;3;4;… etc)

**Column Headings –** identifies each of the columns by a letter (A; B; … ;AA; AB; etc)

**Cell, Active Cell –** Cells are the grid squares of the spreadsheet. The Active Cell or Cells are the cells selected for further manipulation and are indicated by a thicker black line around it

**Worksheet –** The set of grid squares which contain all the basic information. Each worksheet has 256 columns and 65,536 rows.

**Worksheet Tabs –** At the bottom of the page and indicates which Worksheet is being used

Don’t try and memorize all of these definitions at once, but make sure you consult them if you have any doubts. Your Excel screen may vary slightly, but all the basic elements should be there. There are extra toolbars and add-ins which can be used but may not be in the default settings. The mouse cursor usually is an arrow, but will change to an I symbol when over the formula bar, and into a cross when the cursor is over the cells of the spreadsheet.

1. **Data Entry**

Data entry can be undertaken by: 1) importing files from another source (existing files); 2) by direct entry via the screen; or 3) by calculation from existing data (using Excel functions).

1. Existing Files
   1. Select **File** from the Menu Bar followed by **Open** and then the appropriate data source. Excel gives you the option of accessing data from a wide range of sources and formats.
2. Direct Entry
   1. Click on the cell you want to type the data into, and turn it into the Active Cell
   2. Type in your data, which will be shown in the Formula Bar
   3. Complete your data entry by pressing one of the following buttons
      1. Enter Key – enters the data and moves the cursor down one cell
      2. Tab Key – enters data and moves the cursor one cell to the right
      3. Arrow Keys – enters data and moves the cursor in the direction of the arrow
   4. Errors can be erased by using the delete button
   5. Large numbers can be included in any cell, but if they are very large, then they will be shown using the scientific notation (e.g. 663920234534534 turns to 6.6392E+14). This makes no difference to any calculations.

**TASK 1 – Data Entry and Getting Familiar**

1. Log on to the computer and open Excel
2. Type in the data in Table 1 (next page), including column titles
3. **Select a continuous cell range**. Select cell B2 and with the left mouse button held down, drag the cursor to F2. Release the mouse button, and you will cell that these cells have now all been selected.
4. **Select a non-continuous cell range.** Select the range from B2 to B6. Press the **Ctrl** button. With the Ctrl key still pressed, select the range D2 to D6. This will be displayed as shaded on the spreadsheet and indicates the active cells.
5. **Cutting and Pasting.** Select the range of B2 to B8. Click on the cut button (scissors symbol) on the toolbar, or press **Ctrl** + **x.** A flashing border will appear around the active cells. Click on cell D6 and press the paste button or **Ctrl** + v. The active cells will now be pasted in this location.
6. **Moving Cells.** Select a range of cells. Move the cursor down the border of the selected area so that it changes to a cross to a cross with arrows. Drag this selected area right 5 cells and release the button.
7. **Changing Cell Format.** Right click on cell B5 and select the **Format Cells** option. Select the number option and change the format so that the number has 3 decimal points.
8. **Selecting the whole data range.** Click on cell A2. Hold down **Shift** + **Ctrl** + **down arrow.** The whole column should now be selected. This selects all the data in a continuous format in any direction.

RETURN EVERYTHING SO THAT THE TABLE LOOKS THE SAME AND CHANGE THE WORKSHEETS NAME TO COMMUTING TIMES

Table 1: Commuting Times of GRG310C

|  |  |
| --- | --- |
| Transport Mode | Commute Time |
| Car | 15 |
| Car | 12 |
| Walk | 10 |
| Car | 15 |
| Car | 7 |
| Bus | 30 |
| Walk | 10 |
| Car | 15 |
| Car | 15 |
| Bus | 20 |
| Bus | 20 |
| Bike | 7 |
| Bus | 17 |
| Bus | 20 |
| Walk | 10 |
| Car | 8 |
| Walk | 15 |
| Car | 5 |
| Bus | 30 |
| Car | 15 |
| Walk | 15 |
| Bus | 35 |
| Car | 65 |

1. Data Entry via Excel Functions

Data can be entered into an Excel spreadsheet through the use of functions and formulae. Type in the following data into a new Worksheet starting in cell A1.

Table 2: Minutes spent at home, or elsewhere over two weeks

|  |  |  |
| --- | --- | --- |
| Day | Home | Elsewhere |
| 1 | 800 | 640 |
| 2 | 732 | 708 |
| 3 | 912 | 528 |
| 4 | 351 | 1089 |
| 5 | 1091 | 349 |
| 6 | 1301 | 139 |
| 7 | 932 | 508 |
| 8 | 601 | 839 |
| 9 | 1392 | 48 |
| 10 | 984 | 456 |
| 11 | 823 | 617 |
| 12 | 1128 | 312 |
| 13 | 208 | 1232 |
| 14 | 805 | 635 |

Adding a simple formula:

Type Total in cell D1 and press Enter

Type = b2+c2 in cell D2 and press Enter

Extend this formula to other cells:

Click on cell D2 and make it an active cell. There should be a black box around the cell, with a small black square in the bottom right corner. This is called the Fill Handle. Click on the Fill Handle and drag it down to D15 and release the button. You should now have a total value in all the rows.

Adding an Excel function:

Excel has a library of many functions, and it is possible to create your own functions. A function is composed of two parts, a function name and a list of arguments (the numbers or cells which are included in the calculations). Calculate the total minutes spent in a week at home and elsewhere by using the SUM function.

In cell A18 Type: Total

In Cell B18 Type:

=SUM (number 1, number 2, number 3, etc)

Where numbers are cell references, we can click on the cells we wish to include in the calculation.

In Cell B18 re-Type:

=SUM (B2:B15)

Or type =SUM(

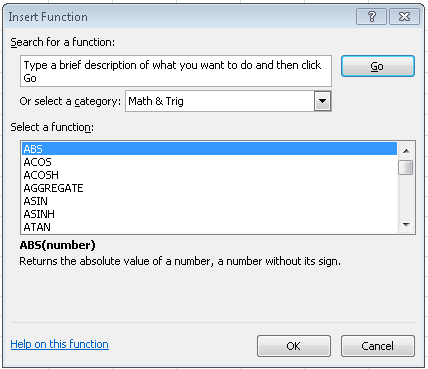
\*and then click on B2 and drag the mouse to B15 and release

Then type ) to close the argument

Do the same for the number of minutes elsewhere and total minutes (C & D columns).

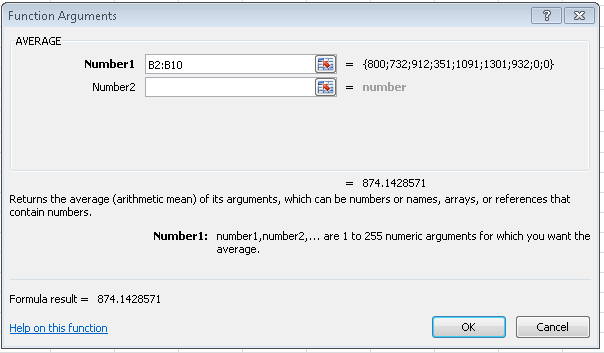
**Function Tab**

In the FORMULA Tab there is a list of all the possible functions that Excel has stored within the library. There is also the ***Insert Function*** button. Navigate to Cell B19, select the cell and then click on the Insert Function button. Click on this. You get the option to search for a function, or select a category for which the function might be stored in. You then get a list, with a description of the function.



We’re interested in knowing the mean minutes someone has spent at home this week. Type in a description of what we’re trying to find, and select the function.

Excel will assume that the cells to average will be all of the cells above B19, this is not true, so use the dialogue box (below) to type in the correct values. You can also click on the dialogue box and drag the mouse over the cells you wish to average. In cell A19, type Mean Minutes.



What are the mean minutes spent elsewhere? Instead or repeating this process, we can click on cell B19, and drag the small black square to C19. This extends the formula across and we have the mean number of minutes spent elsewhere.

Excel uses this relative reference (which is a cell reference based on a range of cells), and so this changes as you drag a function. You can tell Excel to use an absolute reference in a function by using the $ sign.

Type 1440 into cell B21.

In cell A21, type Total minutes.

In cell F1 type “Percentage at home”. Note that the text is too long for the box. Go the column headings, and hover the mouse between A and B. You should get a double sided arrow. Double click this. This should expand the column to fit all the text in.

then in cell F2 type:

=(B2/B$21)\*100

Select F2, and drag the small black square to F15.

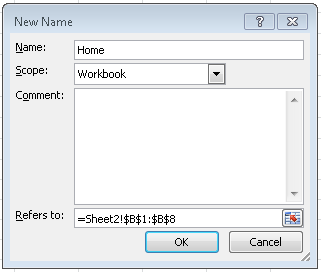
If you click on any of these cells, you will see in the formula tab that B$21 is constant in all of the calculations.

**Repeat this for Percentage Elsewhere using your new skills**

**Creating Range Names**

We can use **Range Names** in order to reference data. This set of cells can be a row or a column, and it would be easier to say “Home” than to continually specify all the cells. To create Range Names:

Select the range B1:B15, right click on the selection and select Define Name.



In an empty cell, redo the analysis for the mean minutes, but type in =AVERAGE(Home), and see if your results are the same.

Give all your columns a Range Name

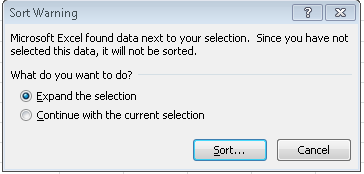
**Sorting and Filtering the Data**

Some common questions with datasets are often aimed at which is the largest or smallest value, how many cases have values above a certain level. We can do this by sorting and filtering the data.

To sort the data:

Select the cell range B1:B15

Click on Data > Sort



This warning will be given. If you do not expand the selection, only that column will be sorted and your data will not be in the correct order anymore.

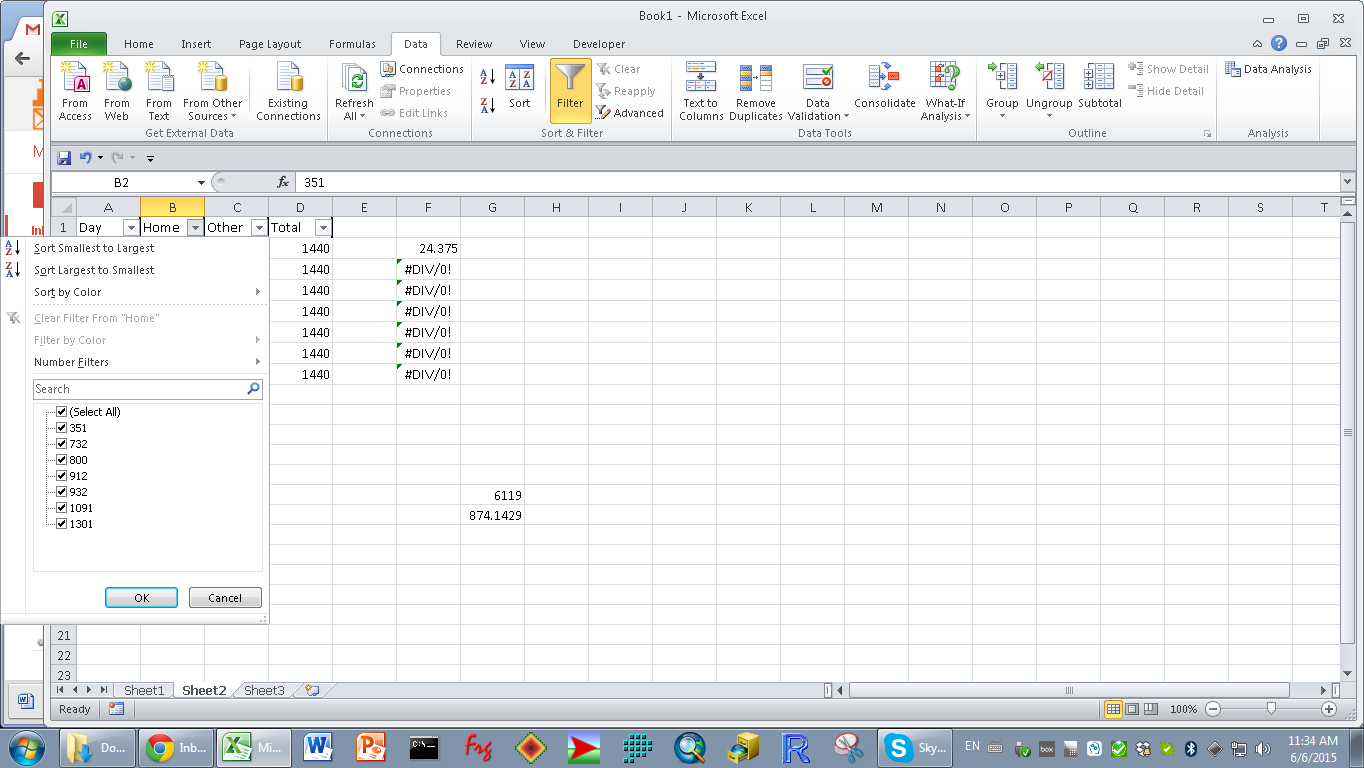
Sort the data by minutes elsewhere, and sort from smallest and largest. We can now easily see which days we spent less time at home.

To filter the data:

To find out what days we spent more than 1000 minutes at home

Click on cell C1, then go to Data > Filter

On the top column, we should have a dropdown menu like this: 



On home, click on Number Filters, and then select the option **Greater Than,** type 1000 into the dialogue and press OK. Now only the days where we spent more than 1000 minutes at home should be displayed. To get out of this option, click on the Filter button again.

You can also add further filters to the option. What happens when you filter by minutes greater than 1000 minutes and less than 200 elsewhere? You will have to perform two separate filtering options. This filtering mechanism works better for two variables which are independent of each other (i.e. profit sales of burgers and hot-dogs at different football game).

**Sorting and Filtering Questions:**

What days did we spend more than 1000 minutes at home?

What days did we spend more than 900 minutes at home and less than 100 minutes elsewhere?

What day did we spend the most time at home?

What day did we spend the least time at home?

**TASK 2 – DESCRIPTIVE STATISTICS**

Return to the commuting time worksheet. Record your answers below.

Using this dataset and the Excel functions, calculate the maximum, minimum, spread (range) of your dataset.

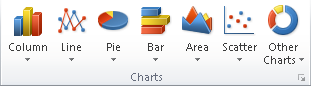
Calculate the mean, median, mode, and standard deviation of your data.

Calculate the skewness, quartiles and any other functions you recognize for your data.

**Creating Graphs and Charts**

One of the main attractions of Excel is that you can create a variety of charts.

Click on the Insert tab, and you have a variety of options.



Play around with displaying the data in various charts. Does every chart make sense for this data?

**Make a pie chart of transport mode**

1. Select the column of transport mode and commute time
2. Go to insert, pie chart, and select one
3. Does our data look correct?? Why not?

A pie chart is also a graphical depiction of frequency, and currently we have a pie chart of commute time. Create a new frequency column for transport mode (i.e. Transport in D1 and Frequency in E1).

**Tip – if you sort transport mode by alphabetical order, it makes it easier to see the frequency. If you also select all the ‘buses’, in the bottom right corner there is a statistic which says ‘Count: 7’. This gives you a quick representation of the frequency.**

Repeat steps 1 and 2 for your new data and then Copy & Paste the Pie chart below.

**TASK 3 – CREATING A HISTOGRAM**

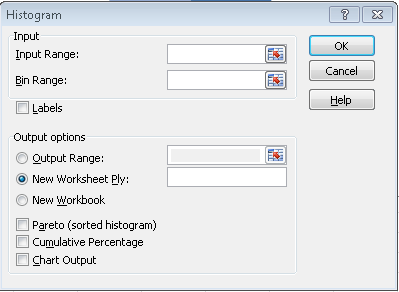
Open a new work sheet. Copy and paste the commuting time column from your first worksheet into the A column.

The next stage is create “bin values” or class intervals into the spreadsheet. Lets take an interval of 5 minutes. In B1 type “Bins”, and then in B2 type 5, B3 type 10. Select both of those cells, and drag the small black square down until you reach 70.

Go to the data tab, and click on Data Analysis

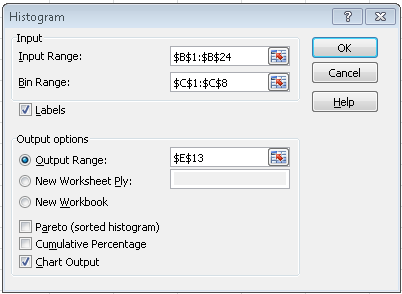
Notice Data Analysis is not there. We need to add an add-in. Click on File > Options. Click on Add-Ins, and then under ‘Manage Add-ins’, Click ‘Go…’. Make sure the Analysis ToolPak is checked, and press OK.

Select Histogram:



Click in the Input Range, and select the commuting time values. Click on the Bin Range and select the Bin Data. If you’ve included the column titles, make sure Labels is selected.

Check Output Range, and choose a cell for the results to be presented. Select Chart Output and press OK.



Copy and paste your results into Word and describe what this shows. What does the histogram show us?

**Extra Credit, Box Plots:**

If you’ve finished in plenty of time, go to <http://www.dummies.com/how-to/content/boxandwhisker-charts-for-excel.html> for instructions on how to make box plots.

Download the Commutes.txt text file from Canvas.

Open a new worksheet, go to file > open. Navigate to where you saved the file. Chose file type as all files. Tick delimited and then next. The text file is tab delineated, make sure this is ticked and press finish.

Follow the instructions one the website, and create 3 box plots.

Note if you are doing this at another time or for other data, Excel doesn’t like making just one box plot, so make sure you either have more than 3 datasets, or trick Excel by creating repeats of your data 3 times (i.e. create three data columns of exactly the same data and create the plots in the same way)