BASICS

MICROSOFT EXCEL
Enter data . Files . Formulas . Functions

Prepared by: Norhaiza Ahmad

Adapted from F. Jolliffe, Kent 2004-2005

Entering data into Excel

Before you start

It is always a good idea to think about how best to organise your data before you start entering it. This is particularly important if you need to export your data from Excel to another package. Many data sets can be arranged as a simple rectangle, called a *data matrix*, with the columns storing different variables (eg. age, height, weight) and each row storing the data for one individual.

File Format

Excel file extensions can be of various types.

XLS, XLSX, .XLSB etc

Use >File >Save As   to know other file extensions for Excel.

Another file format that can be used in Excel is .CSV. CSV is a simple file format used to store tabular data, such as a spreadsheet or database. CSV stands for "comma-separated values". Its data fields are most often separated, or delimited, by a comma. Files in the CSV format can be imported to and exported from programs that store data in tables, such as Microsoft Excel, OpenOffice Calc, Notepad or Google Doc.

TASK 1: Downloading and reading into Excel

At your elearning, go to folder **Datasets** and select the subfolder **Kaggle airbnb**. There are two files in this subfolder. Open and explore both of these files.

The file with extension .csv is a public dataset called: the New York City Airbnb Open Data The dataset consists of an Airbnb listings and metrics in New York City, USA for the year 2019.

Save the file as a Microsoft Excel workbook (file extension .xlsx) into a subfolder (named after you) in partition D of your PC. Name this file **Airbnb**.

Types of data

Each **cell** in a worksheet can store a **single data value**. This might be a number, eg. 4, 17.2 or 2.9E-4, or text, such as Female, Sample 17 or Mary Poppins. If you want a number to be considered as a text value rather than a number, precede it by an apostrophe, e.g. **'2000**. Excel can also deal with dates and times. Consult the on-line help for more information on this. If an entry is too wide to be displayed in the column only part of it will be displayed. Refere to the online Help section on Formatting a Worksheet for details about changing column widths.

Headings

If you are entering different variables in different columns, it is usually sensible to give information about the variables in the first few rows, for example:

	A	B	C
1	Age(years)	Height (m)	Weight (kg)
2	32	1.82	72.6
3	35	1.66	71.0
4	28	17.1	84.1

Entering systematic data

Quite often, your data will have some systematic features – for example, there might be a variable called *Sample Number* that increases by one each time you move down a row. Excel has a useful feature that can speed up data entry in this situation, and which is illustrated by the following example.

TASK 2 (Fill Handles)

Enter the following values in the top left-hand corner of a blank worksheet:

	A	B	C
1	1	1	1
2	1	2	3

Use the mouse to highlight the block of 6 cells. Notice that there is a small square (called the **fill handle**) in the bottom right hand corner. When you move the mouse over the fill handle, the cursor changes to a small '**+**' sign. When this is showing, drag the mouse down a few rows and Excel will automatically fill these with values, as follows:

	A	B	C
1	1	1	1
2	1	2	3
3	1	3	5
4	1	4	7

This illustrates how you can fill cells with a constant value, or with an incrementing sequence. Notice from column C that the increment does not have to be one.

Missing values

Real data sets often contain missing values, and most statistical packages can cope sensibly with missing data. Excel isn't completely satisfactory in this respect. You can leave missing values blank. But if you do this you need to remember that **blank values will treated as zeros in calculations.** Alternatively, you can enter the special missing value indicator **#N/A** into a cell. However, any calculations that involve this cell will generate **#N/A** as the answer.

Correcting errors

If you realise halfway through entering a value that you are entering it in the wrong cell, press the **Esc** key to quit. If you've already finished the entry, you can usually still undo it by clicking on the **Edit** menu and selecting **Undo Typing**. You can delete the contents of a cell with the **Delete** key. You can delete the contents of a block of cells by highlighting the block and then pressing **Delete**. To edit the contents of a cell, just click on the cell and edit the cell contents that are displayed in the formula bar.

Performing Calculations in Excel

There are two basic ways to perform calculations in Excel: **Formulas and Functions**.

Working with Formulas

In Excel, a **formula is an expression** that operates on values in a range of cells or a cell. For example, =A1+A2+A3, which finds the sum of the range of values from cell A1 to Cell A3.

Cells can store formulas that refer to data stored in other cells. When any of the cells used in a formula is updated, the cell containing the result of the calculation is updated automatically.

As well as basic arithmetic operations,

Working with Functions

Functions are predefined formulas in Excel. Excel provides a wide range of specialized functions that enable quite sophisticated calculations to be undertaken.They eliminate laborious manual entry of formulas while giving them human-friendly names. For example: =SUM(A1:A3). The function sums all the values from A1 to A3.

To explore what functions are available, at your Excel Ribbon, choose **Formulas tab**. click on the f_x icon. You will want to look particularly in the category **Math & Trig** and **Statistical.** Do explore other categories such as **Financial** and **Engineering**. You will be familiar with many of the functions already; others will become familiar later in your course.

TASK 3: Extract only the first 20 properties from your Airbnb dataset

Copy the information from the **first 20 properties** of the Airbnb dataset and paste it in a new worksheet of the same file. Name this worksheet AB20

Use this dataset for the following exercises.

Several Fundamental Excel Functions

1. SUM

The SUM function usually aggregates values from a selection of columns or rows from your selected range.

$$=SUM(number1, [number2], …)$$

Suppose we want to calculate the sum of the data values in cells J2 to L21,

1. Click on cell M22 so that it becomes the active cell.

2. Type **=SUM(** (this will appear in the formula bar).

3. With the mouse, click on cell J2, hold the mouse button down, and move the mouse to L21. Then let go of the mouse button. The selected cells will be highlighted by a flashing dashed line border, and the range **J2:L21** will be added to the formula.

4. Click the mouse back in the formula bar and add a closing bracket) to the formula, which will now read **=SUM(J2:L21**), and press **Enter.**

There is an alternative way of selecting a range like **J2:L21** that is particularly useful if the range is too big to fit on the screen and you want to use the mouse for other things, for example to click the arrow buttons on the scroll bar. Instead of holding the mouse button down, click on J2 and press and hold down the **Shift** key. Then move the mouse to D4 and click again. You can select a complete row by clicking on its own row number or a complete column by clicking on its column letter.

Exercise

Use any appropriate empty cell(s) in the worksheet that you have created to write the functions in the following exercises.

=SUM(J2:L2) – A simple selection that sums the values of a row.

=SUM(J2:J8) – A simple selection that sums the values of a column. (note that J-column indicates the prices for the listed air bnb properties)

=SUM(J2:J7, J9, J12:J15) – A sophisticated collection that sums values from range J2 to J7, skips J8, adds J9, jumps J10 and J11, then finally adds from J12 to J15.

=SUM(J2:J8)/20 – Shows you can also turn your function into a formula.

2. AVERAGE

The AVERAGE function is the simple averages of data such as the average prices of prices of the listed air bnb properties.

$$=AVERAGE(number1, [number2], …)$$

Exercise

Use any appropriate empty cell(s) in the worksheet that you have created to write the functions in the following exercises.

=AVERAGE(J2:J11) – Shows a simple average, also similar to (SUM(J2: J11)/10)

3. COUNT

The COUNT function counts all cells in a given range that contain only numeric values.

$$=COUNT(value1, [value2], …)$$

Exercise

Use any appropriate empty cell(s) in the worksheet that you have created to write the functions in the following exercises.

COUNT(A2:P2) – Counts all values that are numerical in the 2_{nd} row.

4. COUNTA

Like the COUNT function, COUNTA counts all cells in a given rage. However, it counts all cells regardless of type. That is, unlike COUNT that only counts numerics, it also counts dates, times, strings, logical values, errors, empty string, or text.

$$=COUNTA(value1, [value2], …)$$

Exercise

Use any appropriate empty cell(s) in the worksheet that you have created to write the functions in the following exercises.

COUNTA(A3:P3) – Counts columns A to P regardless of type.

5. MAX & MIN

The MAX and MIN functions help in finding the maximum number and the minimum number in a range of values.

$$=MIN(number1, [number2], …)$$

Exercise

Find the minimum and maximum airbnb prices

Other summary functions

In addition to the **SUM, MIN** and **MAX** functions, other useful summary functions that are available in Excel include the following:

COUNT	Counts the number of non-empty cells
COUNTBLANK	Counts the number of empty cells
AVERAGE	The arithmetic mean
STDEV,VAR	Measures of variability
SKEW, KURT	skewness and kurtosis
HARMEAN,GEOMEAN, MEDIAN	Other types of average

Beware of empty cells when using these functions- empty cells are treated as if they contained zero. The **SUM** and **COUNT** functions have useful variants, called **SUMIF** and **COUNTIF**.

Exercise

COUNTIF counts data values that satisfy some criterion. For example;

$$COUNTIF(L2:L21,">0")$$

counts how many of the (non-empty) cells in L2:L21 are greater than zero.

Similarly,

$$SUMIF(L2:L21,">0")$$

sums the values in cell L2:L21 that are greater than zero.

SUMIF has a further variant. Column I indicates room type, and column L contain the number of reviews for the properties. The formula

$$SUMIF(I1:I21,"Private room", K1:K21)$$

calculates the total number of reviews for the private rooms in the listings.

Errors in formulas

You will sometimes make typing errors in formulas, for example not matching brackets properly. Excel will report this type of syntax error, sometimes suggesting a suitable correction (which will not always be what you intended!).

Other errors arise when your formula has valid syntax, but produces an error when Excel tries to perform the calculation. Here are some errors and their possible causes:

#####	Not an error. Just means there is no room to display the result in the cell. Make the column wider.
#DIV/0!	Division by zero. If the divisor is a range, it may have an empty cell, which will be treated as zero.
#VALUE!	You have used the wrong type of argument or operand in the formula – e.g. a cell containing text instead of a number.
#NAME?	You may have tried to use a function that doesn't exist, or tried to use a named block of data and got the name wrong.
#REF!	Invalid cell reference. This usually happens when you have moved or copied cells that are involved in the formula.

Next:

Cell references

Use of solver techniques