**Excel Basics & Introduction to Statistics**

This activity aims to introduce you to Microsoft (MS) Excel. We use Excel because it is widely used in many academic and professional settings and is available on both Windows and Mac OSX operating systems.

We will begin at a novice level, assuming that you have minimal experience with MS Excel, data manipulation, statistics, or programming. If you do have experience with these, congratulations… this will be a quick and easy review for you. If you have none and find this pre-activity difficult, please ask your instructor to go over some of the basics with you during class. The lessons will build on each other as you become more comfortable with Excel, and by the end of this course you will have acquired or improved a skill set that includes data management and processing, graphical representation of data, and drawing evidence-based conclusions from real data.

Like any software, MS Excel has strengths, weaknesses and many (sometimes frustrating) idiosyncrasies. Many operations in Excel can be completed a variety of ways. There is a handy reference figure on the window and menus in Excel on the last page of this handout. If you have prior experience with this program and know other ways to complete an operation you are free to do so, however, you should check your approach against the steps laid out here to ensure they produce the same results.

**Learning Objectives**

• Understand the basic organization of an MS Excel sheet

• Learn how to write functions, execute simple operations, and calculate descriptive statistics

**Download the Data:**

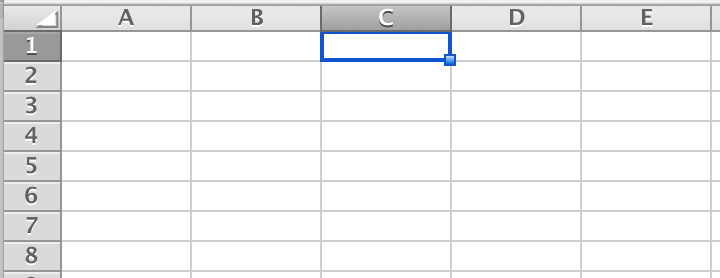
1. Download the example datasheet on the GitHub repository titled “Excel\_Workbook\_Example.xlsx” (<https://github.com/paleodentist/VSA_Session3_Geology>)
2. Open the file.

**Working with the Data:**

**1. Cells and Cell Indices**

The first step in learning to use Excel is learning how to locate data within your **spreadsheet**. The spreadsheet is composed of a grid of **cells**. Each cell can hold **numbers**, **text**, **functions** or other information. The sizes of the cells can be changed, and they can be visually accented with colored highlighting or colored or formatted text.

Along the side of your spreadsheet there are **numbers** increasing in value down the sheet marking **rows**. Along the top **letters** in alphabetical order from left to right identify **columns**. After the 26th column (“Z”), the next column is “AA” then “AB” and so on. In Excel, these numbers and letters identify unique cells (letter + number) that tell Excel which cells you would like to operate on. These identifiers are called **cell indices** (plural of index = “indices”). Every single cell on an Excel spreadsheet has a **cell index**. The letter always comes first, indicating which column the cell is located. The numeral comes second (no spaces) and it indicates the row in which the cell is located. Thus, the cell in the top left-hand corner of the spreadsheet has the cell index, A1.

Excel helps you locate the cell indices by shading the row and column indices for the selected cell. If you were to click on the first cell in the 3rd column you would notice that both ‘C’ and ‘1’ are shaded (**see picture at left**). Likewise if you select more than one cell, the selected indices associated with the selected cells will also receive the same treatment.

**Question 1:** **What are the cell indices that are highlighted in yellow on the example spreadsheet?**

**2. Vectors and Arrays: Selecting multiple data points**

You will notice that under the highlighted cell in column A there are five numbers. When you check their indices, you will notice that they are: A7, A8, A9, A10, and A11 (in Excel syntax we could also write this as: A7:A11). This is an example of a **vector**, with the **dimensions** of 5 rows and 1 column, or 5 x 1.

If you were to copy and paste the same vector immediately to the right in column B, you would now have an **array** with the dimensions of 5 rows and 2 columns, or 5 x 2. For our purposes in the first few lessons, we will not delve too deeply within this topic; however, we will explore vectors and arrays in later lessons.

**Question 2: What are the dimensions for the array highlighted in green?**

**3. Excel Functions and Operations**

Excel can do more than just store data in vectors and arrays; through Excel you can perform all of the **operations** that you can do on your calculator and more. In this section we will explore the basics of writing **functions** in Excel.

**The Importance of the Equals “=” Sign**

For all operations that we will cover in this course you will start the operation with an **equals sign** ‘=’ within an empty cell to tell Excel that you are performing a mathematical or functional operation.

Let's walk through an example together. Perform the following in the Excel Basics worksheet:

1. Select cell G6 and enter the value 10, and in cell H6 enter the value 5 (no equals sign required yet).
2. Now, in cell I6 we will represent the sum of cells G6 and H6. Enter an equals sign and then write G6 + H6 into cell I6. Press enter. Voila, you have computed the sum of two different cells.
3. Click on cell I6. Notice that your equation appears in the large box above the columns of your spreadsheet (this box is called the **formula bar**). Go into that box, and delete everything but the ‘=’ sign. Now, you will create the same equation by clicking on the G6 cell, then typing the ‘+’ sign, then clicking on the H6 cell and pressing enter. This is usually more efficient than typing your equations manually.

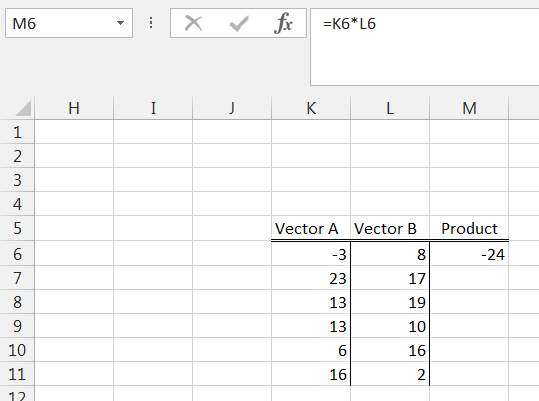
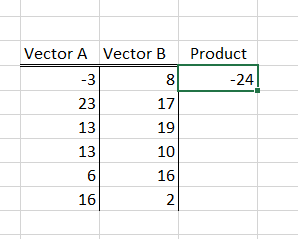
Now let's try something more complicated.

1. Locate Vector A, Vector B, and the vector entitled “Product" (in columns K, L, and M). Your objective will be to find the product of each pair of values located in vectors A and B. You
2. To begin the operation, enter an equals sign into cell M6.
3. Next, click on cell K6, type the ‘\*’ symbol for multiplication, then click on the L6 cell. Press enter.
4. You will want to repeat this process for the remaining pairs until you have filled out the Product vector.

**Question 3: What is the value located in cell M11?**

**Protip 1:** After you have completed these steps, delete the contents of cells M7:M11 (but not M6!) If you click on M6, you will see the function displayed in the large menu box above. Notice that when the cell is highlighted, there is a small square on the bottom right of the cell (**see picture below**). Click and hold that small square, and drag it down to highlight all the cells you want to fill with the product equation. What happens?

Place your cursor on this square. When your cursor turns into a ‘+’ you can click, hold down the button, and drag down to copy the equation to the highlighted cells.



**Protip 2:** Again, delete the contents of M7:M11 and select M6 to make the small square appear at the lower right corner of the cell again. Quickly double-click the little square. What happens?

Mastering these types of Excel shortcuts will save you enormous amounts of time and grief. We will highlight them in assignments as “**Protips”** along with other tricks, hacks and hints for managing and analyzing data in smart and efficient ways.

**Functions in Excel**

Lastly, let's take a look at using specific functions within Excel by computing the **mean (the average), standard deviation, median, maximum,** and **minimum** of the values in the Product vector. Explanations of these terms are on the following page.

1. Enter an equals sign into cell O6
2. Next, execute the Excel function AVERAGE() by typing AVERAGE to the right of the equals sign.
3. Now that the function has been called, it is time to enter the argument. First, type a right-facing parenthetical, ‘(‘.
4. Now you need to specify which cells you want to average. You could do this manually, but in most cases it is easier to click and highlight the cells you want to perform an operation on.
5. Since we are interested in finding the mean value of ‘Product,’ you simply select all six cells in the vector (M6:M11)
6. Next, enclose the argument within the function using a left-facing parenthetical ‘)’. Your function bar should read like this:

=AVERAGE(M6:M11)

1. Hit ‘Enter.’ You will now have the mean product in cell O6.

**Note** that the calculated value from a function will automatically change if the contents of the source cells are changed. To see this for yourself, try deleting the contents of cell M6 and watch what happens to the number displayed in cell O6. Recalculate the value for M6 before proceeding. Be careful when working with spreadsheets that use functions linked to cells within the spreadsheet, one small accidental mistake can have cascading effects.

1. Do the same thing with the remaining statistics (cells P6 through S6).

**Hint:** If you get stuck, click on the ‘ƒx’ symbol next to the formula bar and find the function you need from the dropdown menu. Consult the “Anatomy of an Excel Window” at the end of this activity for more help finding this and other features mentioned in the activity.

**Question 4: Check that your calculations make sense. The “Maximum” and “Minimum” should be the largest and smallest numbers in the vector. The “Mean” and “Median” should be values between this. Include the mean, standard deviation, median, minimum, and maximum in your HW document.**

**Descriptive Statistics**

The values you entered in cells O6 through S6 are called descriptive statistics. The primary goal of descriptive statistics is to describe basic features of the data that tell you something about the dataset as a whole, and allow comparison with other datasets.

• **data**—numbers or measurements that are observed and collected

• **variables**—any characteristics, numbers, or quantities that can be described, measured, or counted; a variable is essentially an item of data. In a typical Excel spreadsheet, columns represent different variables and rows represent different recorded observations of those variables.

• **mean**—this is the “average” of a set of numbers, attained by summing the all of the numbers and dividing the total by the number of numbers. (Excel function: AVERAGE)

• **median**—this is the middlemost value when a set of numbers is listed in order, e.g., of [1, 3, 4, 5, 8], 4 is the median (Excel function: MEDIAN)

• **variance**—how far a set of numbers is spread out, i.e., the average of the squared differences from the mean. (Excel function: VAR.P)

• **standard deviation**—another measure of how spread out the numbers are, this is the square root of the variance. (Excel function: STDEV.P). \*Note that Excel has functions for other versions of the standard deviation; we will discuss the differences between these in a later activity.

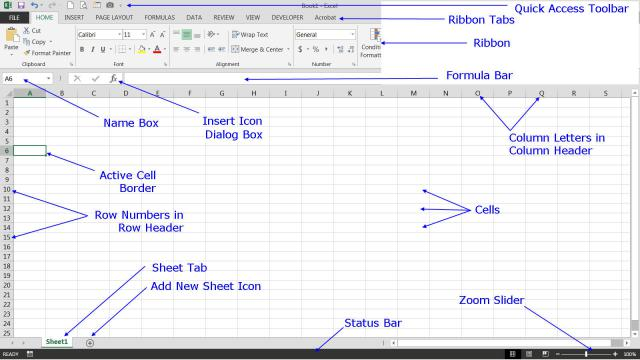
• **maximum**—the largest number in a set (Excel function: MAX)

• **minimum**—the smallest number in a set (Excel function: MIN)

**Statistics Resources**

* Intro to basic statistics terminology: <http://bobhall.tamu.edu/FiniteMath/Module8/Introduction.html>
* Useful Excel functions:

<http://sciencefair.math.iit.edu/analysis/statistics/>



Anatomy of an Excel Window Photo Credit: © Ted French

Source: <http://spreadsheets.about.com/od/excel101/ss/2014-08-16-Parts-Of-The-Excel-2013-Screen.htm>

Note that the appearance might vary somewhat depending on your operating system and which version of Office you are using. However, you should be able to find these basic features in more or less the same location.