

# HSC3103 – RESEARCH DESIGN AND EPIDEMIOLOGY

## TUTORIAL 8

### CORRELATION AND REGRESSION

#### LEARNING OUTCOMES

- Understand applications of common statistical tests
- Determine the most appropriate analysis for your data
- Learn how to perform statistical tests within Microsoft Excel

#### OVERVIEW

The purpose of this tutorial help you identify and perform the most appropriate statistical test for your project data. We will be examining three types of tests over Weeks 7 to 9:

- Week 7: T-tests and ANOVAs
- Week 8: Correlation and Regression
- Week 9: Non-parametric tests

Each week we will examine the theory behind each test in detail, however, if you have determined you require a different test to the topic of the week, the resources can be made available to you if required. The more commonly required tests are covered in earlier weeks to minimise this being necessary however.

#### **PART 1: Identifying an Appropriate Statistical Test**

##### **Purpose**

Collected data alone cannot answer a research question. As seen in Week 6, data must be managed, analysed and finally interpreted. As correct tools are required to measure data, the correct test must be selected to analyse data. This week we will identify whether a correlation or linear regression is appropriate for your data.

##### **Activity 1: Identifying an Appropriate Statistical Test**

Based on your experimental design and types of data collected, examine whether a correlation or regression would be appropriate. If not, speak to your tutor about accessing the tutorial notes Week 9 to discuss some of the more specialised tests. In either case, write your justification for or against a correlation or regression below:

## PART 2: Performing Correlation and Regression in Excel

### Purpose

As we examined in the previous tutorial, Microsoft Excel is a common spreadsheet program which can be used to perform a variety of calculations, create a range of graphs and perform basic analyses.

### Activity 1 – Customising Excel

Excel's basic functions are useful, but for more involved statistical analyses you will require one of the many available add-ins for Excel. These can be accessed via the following (see Figures 1-5):

**File → Options → Add-ins → Manage: → Excel Add-ins → Go → Analysis ToolPak → OK**

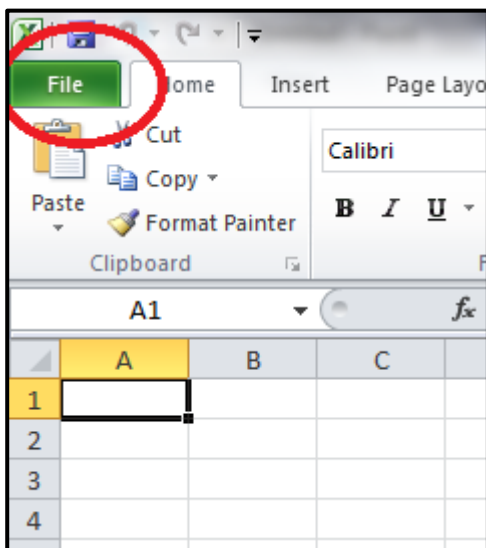


Figure 1: Locating the 'File' Tab

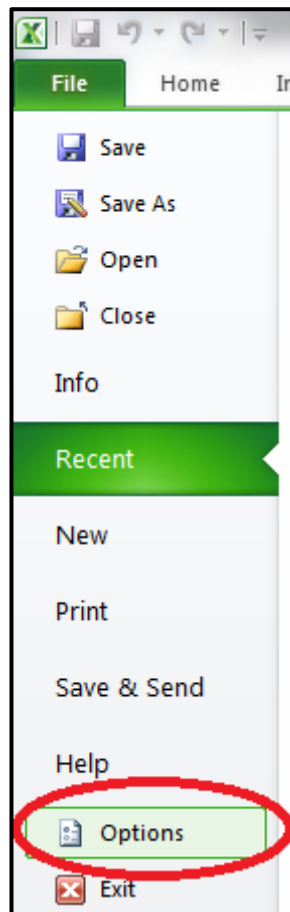


Figure 2: Locating the 'Options' window

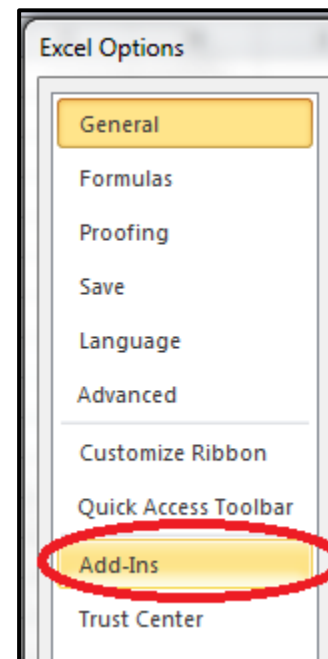
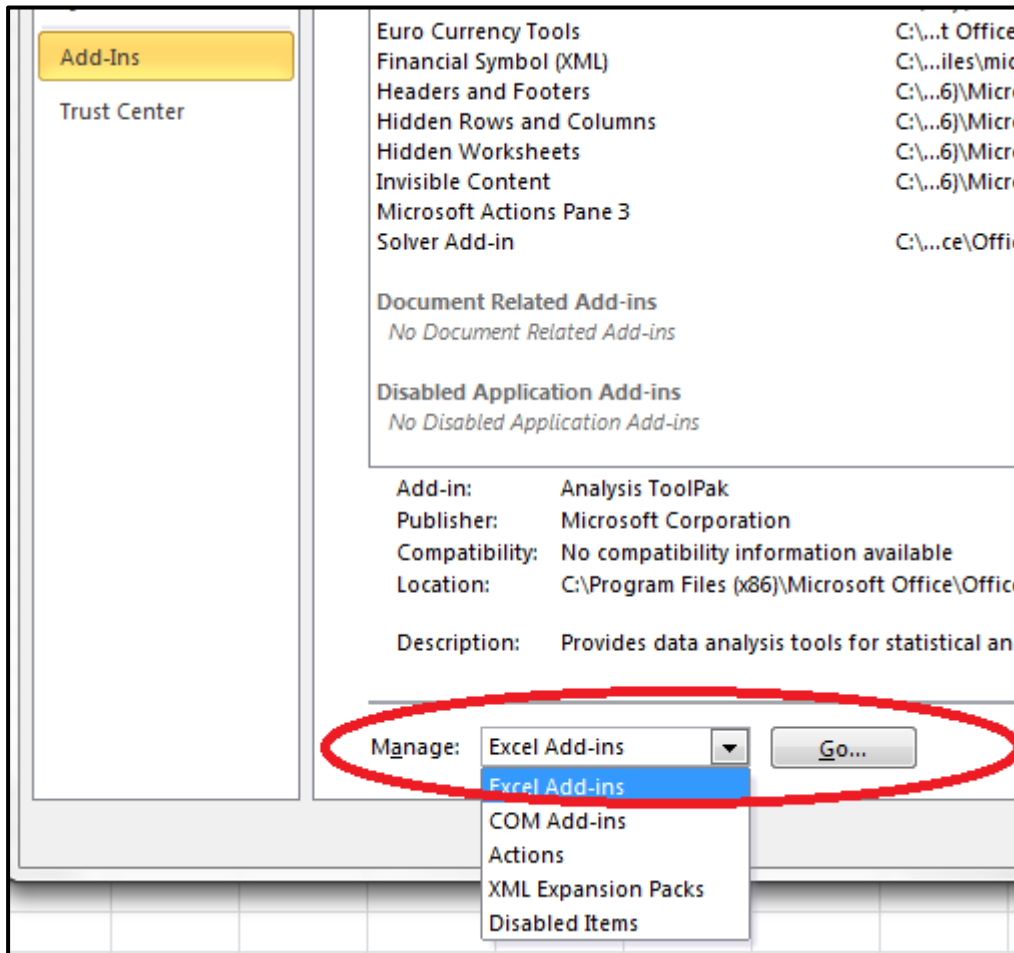
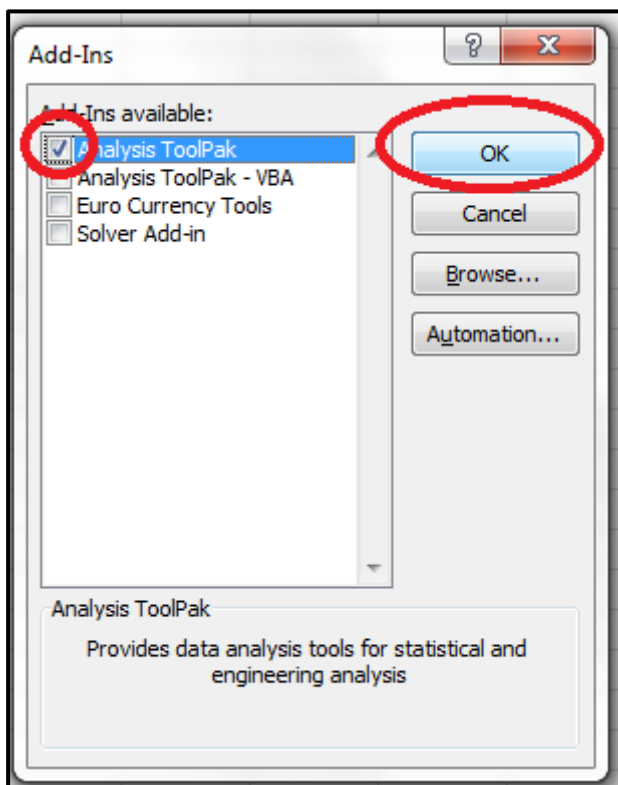


Figure 3: Locating the 'Add-ins' tab



**Figure 4:** Selecting add-ins to manage



**Figure 5:** Installing Analysis ToolPak

## Activity 2 – Using the Analysis ToolPak in Excel

With the Analysis ToolPak installed we can now perform t-tests and ANOVAs. Under the ‘Data’ tab (see Figure 6) in Excel you will now see a new option on the far right which is labelled ‘Analysis’ (see Figure 7).

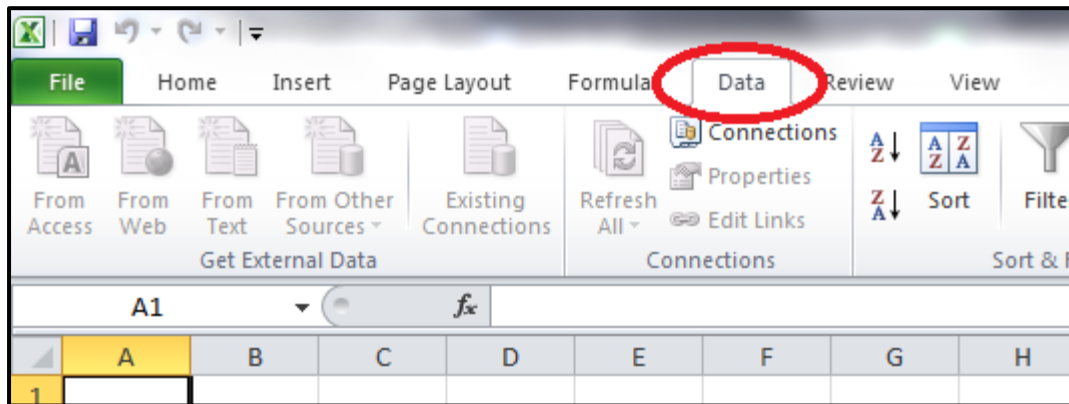


Figure 6: Locating the ‘Data’ tab

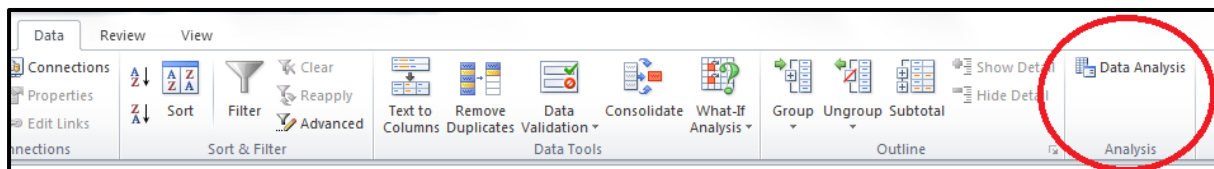


Figure 7: Opening the ‘Data Analysis’ window

Selecting Data Analysis will bring up a new window (see Figure 8) in which you can select which type of test you would like to perform. For this example we will select “Regression”.

*Note:* Actual selected test will vary by what is appropriately for your data.

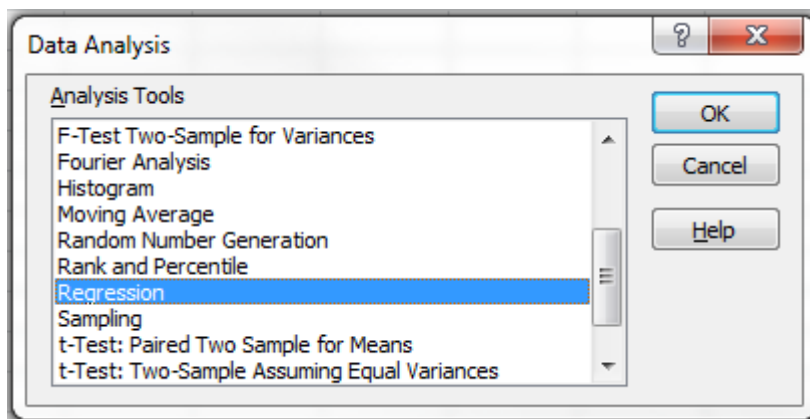
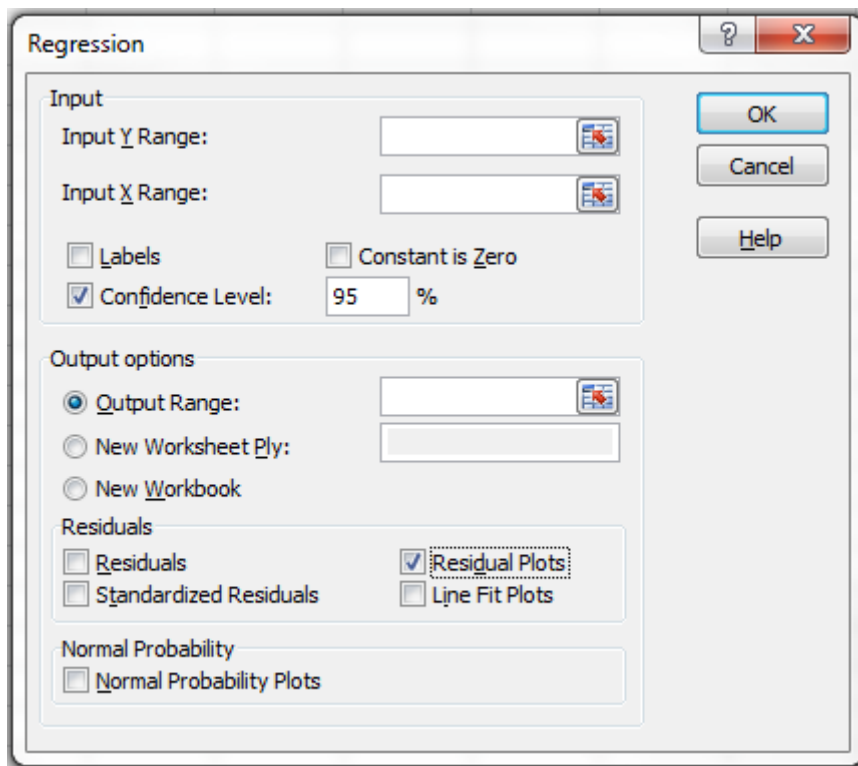


Figure 8: Data Analysis window

Selecting this test will open other window (see Figure 9) which will ask for various input data, required alpha value, and where you would like the results to be output (e.g. specific cells on current sheet, new sheet, or new workbook).



**Figure 9:** Options for Regression (with confidence level and residual plot options enabled)

With the correct information, the test will be performed and the results will appear as selected. Additional options and requirements will depend on the chosen test.

Your tutor can assist you with specific options and data layout appropriate for each test, as well as assistance in interpreting output data as needed.

## Activity 2 – Alternate Programs

Microsoft Excel is a retail program, however, free alternatives do exist which share similar functions and even similar (or identical) commands. Issues may arise due to incompatible file types between software packages. Options include:

- Apache OpenOffice
  - Open Source office productivity software
  - Excel equivalent is “Calc”
  - <https://www.openoffice.org/>
  
- LibreOffice
  - Open Source office productivity software
  - Excel equivalent is “Calc”
  - <https://www.libreoffice.org/>
  
- Google Sheets
  - Cloud-based office productivity software
  - Requires Google account to edit files
  - <https://www.google.com/sheets/about/>

Another non-spreadsheet option is **VassarStats**; a website which performs a variety of statistical analyses programmed in JavaScript and has an accompanying online textbook for all tests offered.

- <http://vassarstats.net/>

Please note this is not an official Vassar College website, but the name is used with permission; the website’s author (Dr Richard Lowry) is an Emeritus Professor in Psychology who was based at Vassar College at the time of its development.

## REFERENCE LIST

Stewart, A. (2016), *Basic Statistics and Epidemiology: A Practical Guide* (4th Edition) (pp 65-82). Boca Raton, FL: CRC Press