INSTRUCTIONS

# Download:

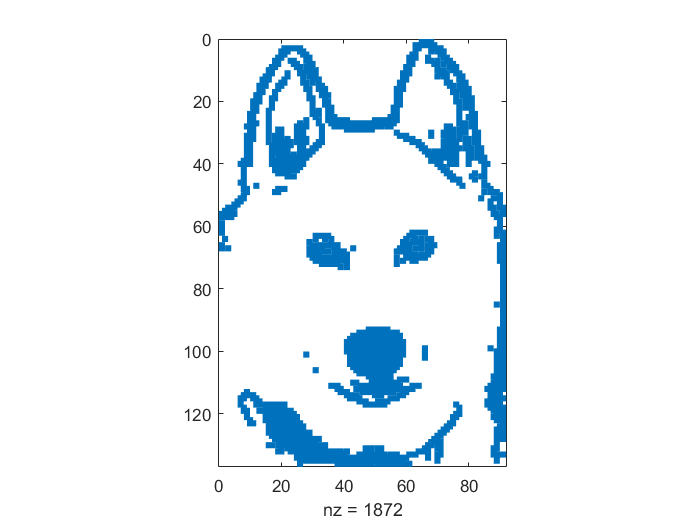
1. Download either the doge or anova folder (should be a zipped folder after downloading)
2. Extract all files

# Install:

1. Go to folder “for redistribution” and double click the .exe
2. Follow the prompts to install the function as well as MATLAB runtime
3. Open Microsoft Excel
4. Go to “File”--> “Options”--> “add ins”-->Manage Excel add in--> “Go”-->”Browse”-->navigate to the folder “for\_redistribution\_files\_only” and select “anova.xla”

# Use the doge formula:

1. Open Microsoft Excel
2. In an empty cell, type “=doge()” (the doge function has no argument)
3. This brings up a doge. Give it a minute, the first time the function calls it takes a little time, using it again should be faster.



# Use the anova formula:

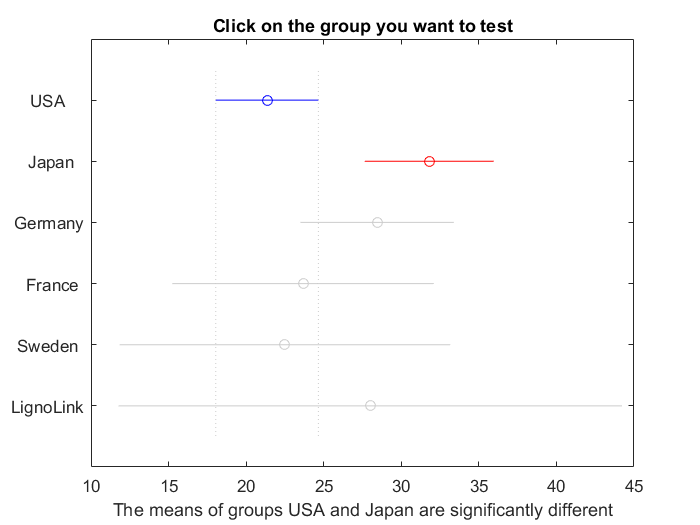
1. In an empty cell, type “=anova(x),” where “x” is the data that you want to conduct a one way analysis of variance on.
2. The data that is highlighted should be in the following format:

|  |  |
| --- | --- |
| name | value |
| name | value |
| name | value |

For instance, try highlighting the following data (car origins and their MPGs):

|  |  |
| --- | --- |
| USA | 18 |
| USA | 15 |
| USA | 18 |
| USA | 16 |
| USA | 17 |
| USA | 15 |
| USA | 14 |
| USA | 14 |
| USA | 14 |
| USA | 15 |
| USA | 15 |
| USA | 14 |
| Japan | 24 |
| USA | 22 |
| USA | 18 |
| USA | 21 |
| Japan | 27 |
| Germany | 26 |
| France | 25 |
| Germany | 24 |
| Sweden | 25 |
| Germany | 26 |
| USA | 21 |
| USA | 10 |
| USA | 10 |
| USA | 11 |
| USA | 9 |
| LignoLink | 28 |
| Germany | 25 |
| USA | 25 |
| USA | 26 |
| France | 27 |
| USA | 17.5 |
| USA | 16 |
| USA | 15.5 |
| USA | 14.5 |
| USA | 22 |
| USA | 22 |
| USA | 24 |
| USA | 22.5 |
| USA | 29 |
| USA | 24.5 |
| Germany | 29 |
| Japan | 33 |
| USA | 20 |
| USA | 18 |
| USA | 18.5 |
| USA | 17.5 |
| Germany | 29.5 |
| Japan | 32 |
| Japan | 28 |
| USA | 26.5 |
| Sweden | 20 |
| USA | 13 |
| France | 19 |
| Japan | 19 |
| Germany | 16.5 |
| USA | 16.5 |
| USA | 13 |
| USA | 13 |
| USA | 13 |
| USA | 28 |
| USA | 27 |
| USA | 34 |
| USA | 31 |
| USA | 29 |
| USA | 27 |
| USA | 24 |
| USA | 23 |
| Germany | 36 |
| Japan | 37 |
| Japan | 31 |
| USA | 38 |
| USA | 36 |
| Japan | 36 |
| Japan | 36 |
| Japan | 34 |
| Japan | 38 |
| Japan | 32 |
| Japan | 38 |
| USA | 25 |
| USA | 38 |
| USA | 26 |
| USA | 22 |
| Japan | 32 |
| USA | 36 |
| USA | 27 |
| USA | 27 |
| Germany | 44 |
| USA | 32 |
| USA | 28 |
| USA | 31 |

The anova function returns the following figure:

  
Try selecting different groups to test the means for statistically significant differences!

## Details:

1. The anova function tests for statistically significant differences between means at the significance level of 95% (alpha = 0.05)
2. The critical value is determined using the Tukey-Kramer honest significant difference criterion
3. The guts of the function are taken from this documentation: <http://www.mathworks.com/help/stats/multcompare.html?refresh=true>
4. The function ignores any missing rows. If there is an empty cell in the leftmost column (the names column) then the data in the cell to the right will be ignored. Likewise, if there is a name with no data point, this row will also be ignored. Basically, empty rows do not trip up the anova function and can be included in the highlighted block. I did this so that blocks of different groups can be separated in the Excel spreadsheet with empty rows to make it easy on the eyes.
5. The function ignores any columns in between the left and rightmost columns that are in the highlighted block. This means that if the following block was highlighted:

|  |  |  |
| --- | --- | --- |
| group | Data type 1 | Data type 2 |
| goup | Data type 1 | Data type 2 |

Then the test would conduct a one way anova on the groups (leftmost column) for Data type 2 (rightmost column). If you want to run it on Data type 1, simply highlight:

|  |  |
| --- | --- |
| group | Data type 1 |
| goup | Data type 1 |

The reason for this is because in spreadsheets I often have data in the form of:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Data 1 | Data 2 | Data 3 | Data 4 |

For instance, maybe something like this:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Company name** | **return** | **sales** | **income** | **tax** |
| McCormick | 10 | 5 | 2.2 | 1.0993 |
| McCormick | 11 | 5.1 | 2.2 | 1.0994 |
|  |  |  |  |  |
| Ocean Spray | 11 | 4 | 2.21 | 1.1100 |
| Ocean Spray | 11 | 4.1 | 2.20 | 1.0990 |

The way the anova function is set up, you can easily conduct the test on any of the metrics (return, sales, income, or tax) just by highlighting different regions in the spreadsheet. If you want to do income, highlight this part (don’t highlight the table headings)

|  |  |  |  |
| --- | --- | --- | --- |
| McCormick | 10 | 5 | 2.2 |
| McCormick | 11 | 5.1 | 2.2 |
|  |  |  |  |
| Ocean Spray | 11 | 4 | 2.21 |
| Ocean Spray | 11 | 4.1 | 2.20 |

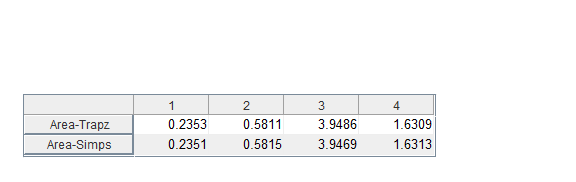
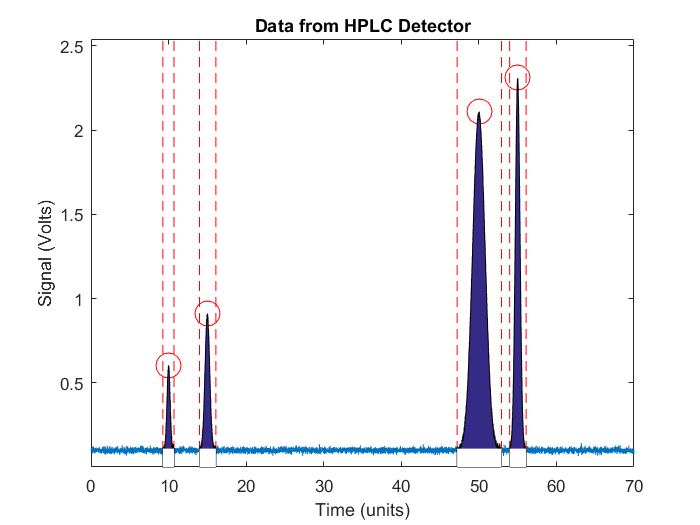
Only the company name and rightmost column (income) will be used in the test, and the blank row will be ignored.

# Use the signalint function:

* + - 1. Get some data where you have a x variable (like time) and a y value that is some signal (like voltage from an instrument) where you might want to find the area under the peaks in the signal. Quantitative chromatography (like HPLC) is an example.
      2. Import the data into excel with time on in the left column and signal (voltage) in the right column, like so (may be thousands of cells of data):

|  |  |
| --- | --- |
| Time(seconds) | Voltage(millivolts) |
| 0.1 | 0.05 |

* + - 1. Call the formula “=signalint(data,trigger)” where “data” is the entire data frame of times and voltage, and trigger is a single number, the “trigger” height for a peak. Do not highlight the headers. Any peaks must have this as a minimum of both *height* and *prominence* to be automatically detected. For example, y could be 0.5. For this signal, this gives you the following:

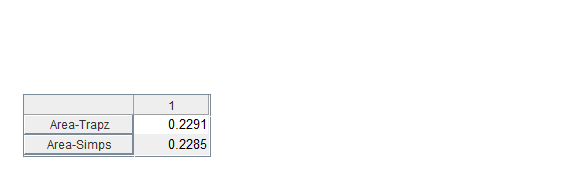
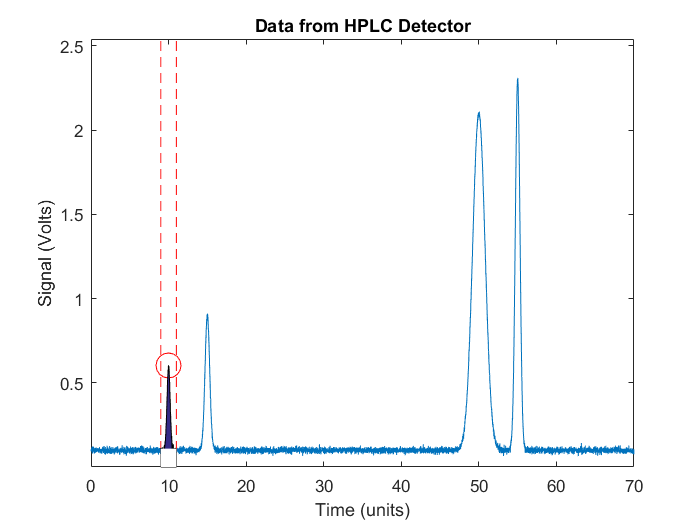


The plot shows the trace of the signal, and the peaks that were found as well as shades the area that is calculated. The table reports the area as found through Simpson’s method and the trapezoidal method.

# Use manualint

* + - 1. Same as signalint except without automatic peak detection, and only one peak calculated at a time. The syntax is “=manualint(data,trigger,peakstart,peakend)”, where data and trigger are the same for signalint, and the peakstart and peakend values are the x values where you want to start and end. This allows you to manually adjust if signalint picks up something weird for the start and end of a peak.

Example shown (same data as previous) but by calling “=manualint(data,0.5,9,11)”



Only one peak is found and the area is reported.

Tip: Try running signalint first and if it needs to be adjusted for specific peaks, run manualint. This way you can avoid extra work if signalint works well without modification!