2016

Correlation Analysis with Excel Handout

Kelly A. Rodgers
CUNY Borough of Manhattan Community College

CORRELATION ANALYSIS WITH EXCEL

Downloading data from Google Sheets:

1. Go to sheets.google.com
2. Select and open the appropriate sheet.
3. Go to File → Download As → Microsoft Excel
4. Find where the file was saved and click on it to open it in Excel.

Downloading Data Analysis ToolPak

1. If you click on 'Data' and do not see 'Data Analysis' then you need to download the Data Analysis Toolpak. To do so:
    a. Go to File → Options → Add-ins → On the screen there's "Manage" and a drop down menu. Click "Go…"
    b. Check the boxes next to 'Analysis ToolPak' and 'Analysis ToolPak –VBA'
    c. Click ok.
    d. Now, when you go to Data → You should see "Data Analysis" as an option.

Creating a correlation matrix in Excel

1. First thing – ALL of the data must be in number format. So if you have data that is not (for example, if you have 'yes' or 'no' as responses), you must code these into numbers, such as 1 = Yes and 2 = No.

2. Go to Data → Data Analysis → Correlation → OK
3. Click on small graphic next to "input range."
4. Select all of the data that you would like to investigate correlations by highlighting it.
5. Click on the small graphic again. The window opens back up.
6. If you have selected your column labels too, check "Labels in First Row" so Excel knows that there's no data there, just titles.
7. Click OK.

Creating a Scatterplot in Excel

You may want to show your strong correlations in the form of a scatterplot. To do so,

1. Highlight the two columns with the variables you would like to use. If they are not next to each other, select one, and then hold down your CONTROL key and select the other and they should both be highlighted.
2. Go to Insert → Scatter → Scatter with only markers.

Creating a frequency histogram in Excel

1. Insert three columns to the right of the column that you want do calculate the frequency. For example, if you want to find out the frequency of the ages of the people in your study, find your age column in your data and insert three columns to the right of it.
2. Label one of the new columns "Bins" label one "Frequency" and the third one "Intervals".

3. Under the “Interval” column, decide what intervals would you like to break your data up into. For example, if we’re doing ages of college students, then I might want to have intervals of 3 years: 18-20, 21-23, 24-26, 27-29, 30+. There’s no need to go below 18 because most people will be at least 18.

   Enter your intervals just as listed above in the “Interval” column, one interval per line.
4. Under “Bins” enter the last number in each interval. So in our age example, you would put 20, 23, 26, 29. Do NOT include your last number here, so don’t include the 30+.
5. Highlight all of the cells underneath your Frequency title. Highlight all the way down to one cell more than your bins. So in the age example, I have 4 bins (see #4 above), so I need to highlight 5 cells underneath my title “Frequency”.

   In the function box (marked by ***fx***), type: =FREQUENCY(data array,bin array)

   That “data array” is just the range of cells containing the data that we want to do frequencies of. So, it’s the range of cells that our ages are in, in the format First cell:Last cell

   The Bin array is the range of cells in the bin column, also in the format First cell:last cell.
6. Hit CONTROL + SHIFT + ENTER all at once. Your frequencies appear. (on Mac, SHIFT+ CONTROL + ENTER)
7. Highlight all of your frequencies in the column.
8. INSERT → choose a chart (we usually will use a simple bar graph here). Chart appears with your frequencies.
9. Your chart will just number all of your data as 1, 2, 3, 4, etc. If you want your actual ranges to show (in our example, if you want to see something like 18-20, 21-23, etc. rather than just 1, 2, 3…)
   a. DESIGN → Select Data → under Horizontal axis, click EDIT.
   b. Drag and select the intervals you have under “Intervals”
   c. Click OK.