

For Research Methods CSJ – Input from the Students of 2010
(Last update – March 23, 2010)
General Background Information

Here are some links for Excel tutorials:

<http://www.youtube.com/watch?v=OTz2PQ-CdJU>

You tube introduction to Excel – you will note that there are other You Tube instructional videos.

<http://psychology.georgetown.edu/resources/researchmethods/computer/8501.html>

Nice place to start with a look at the basics

<http://cameron.econ.ucdavis.edu/excel/ex11histogram.html>

Nice tutorial on making a histogram (bar chart)

<http://phoenix.phys.clemson.edu/tutorials/excel/graph.html>

Sections 7-11

http://www.businessbookmall.com/Statistics%20Internet%20Library.htm#Doing_Statistics_With_Excel

Powerpoint presentations of most of the major statistical concepts

<http://cs.marlboro.edu/courses/spring2008/statistics/wiki/wiki.attachments/993938.pdf>

The above document is the most comprehensive on line free resource for Excel 2007 Statistics

In addition the information below may be helpful.

Getting Started (Most of this material was taken from Microsoft Excel Help)

Open excel. Click on Tools. On the drop down menu click “Add-ins.” Check “Analysis ToolPak” and click O.K. The next time you click on Tools, a “Data Analysis” option should appear. Click it. Select the appropriate tool and click on it. A window will open asking for information on data ranges, and other parameters.

About statistical analysis tools

Microsoft Excel provides a set of data analysis tools — called the Analysis ToolPak — that you can use to save steps when you develop complex statistical or engineering analyses. You provide the data and parameters for each analysis; the tool uses the appropriate statistical or engineering macro functions and then displays the results in an output table. Some tools generate charts in addition to output tables.

Related worksheet functions Excel provides many other statistical, financial, and engineering worksheet functions. Some of the statistical functions are built-in and others become available when you install the Analysis ToolPak.

Accessing the data analysis tools The Analysis ToolPak includes the tools described below. To access these tools, click Data Analysis on the Tools menu. If the Data Analysis command is not available, you need to load the Analysis ToolPak [add-in](#) program.

[Anova](#)

The Anova analysis tools provide different types of variance analysis. The tool to use depends on the number of factors and the number of samples you have from the populations you want to test.

Anova: Single Factor This tool performs a simple analysis of variance, testing the hypothesis that means from two or more samples are equal (drawn from populations with the same mean). This technique expands on the tests for two means, such as the t-test.

Anova: Two-Factor With Replication This analysis tool performs an extension of the single-factor anova that includes more than one sample for each group of data.

Anova: Two-Factor Without Replication This analysis tool performs a two-factor anova that does not include more than one sampling per group, testing the hypothesis that means from two or more samples are equal (drawn from populations with the same mean). This technique expands on tests for two means, such as the t-test.

Correlation

The Correlation analysis tool measures the relationship between two data sets that are scaled to be independent of the unit of measurement. The population correlation calculation returns the covariance of two data sets divided by the product of their standard deviations based on the following formulas.

You can use the correlation analysis tool to determine whether two ranges of data move together — that is, whether large values of one set are associated with large values of the other (positive correlation), whether small values of one set are associated with large values of the other (negative correlation), or whether values in both sets are unrelated (correlation near zero).

Descriptive Statistics

The Descriptive Statistics analysis tool generates a report of univariate statistics for data in the input range, providing information about the central tendency and variability of your data.

Chi Square – See the hand out.

The Data Set Assignment

Part 1 – Qualitative Analysis

Every student needs to do one qualitative analysis, and every group needs to do an analysis of the “Define Compassion” question. When looking at all the answers given for the “Define Compassion” question it is best if you copy the data column for that answer, move it into its own Excel sheet, and then expand the cell size so that you can read all of the responses. What you will want to do is to come up with themes, or categories, into which you will place the responses. After that you may want to code the responses according to your category definitions. You will need also to provide a rationale, and some examples, for these categories. You can have a frequency bar graph, or simply state the frequencies. Also read the “Theory of Compassion Development” paper (see link: http://compassionspace.com/Papers_on_Basic_Compassion.html) and discuss the relationship between the theory and the data. Is there any other data in the survey that might support/refute the theory?

Since each group has multiple members, you will have multiple qualitative analyses done. For each qualitative analysis, be sure you discuss themes, rationale, frequency and links to either (or both) the qualitative definition question or the Theory of Compassion Development.

Part 2 – Descriptive Statistics

Frequency

1. Create a new sheet with the data for which you would like to calculate the frequency. Cut and paste the cells for the survey question you are considering, including headings so you know what information is on this sheet.
2. For questions with multiple responses, (i.e. years of professional experience where the answers are 1 to 5, 6 to 10, etc.), put all of the answers in the first column (farthest left).
3. Create a separate table like the one shown below for your results.
4. Before you choose this function, highlight the cell in which you would like the results to display.
5. Despite common sense, you are going to use the function COUNTIF for frequencies with text, or *summing contents by certain criteria*.
6. Go to function. Select COUNTIF.
7. For Range, indicate the entire left column by a) selecting the arrow and manually choosing from the sheet of data or b) indicating the range by typing it (i.e. A3:A155).
8. For Criteria, select, as described above, the criteria that you would like it to count (i.e. less than 1).
9. Complete a COUNTIF for each of the criteria. With the frequency for each criteria, a graph can be made for the frequency.

Example Table

Criteria (Yrs Prof Exp)	Frequency
Less than 1	(highlight this cell)
1 to 5	
6 to 10	
11 to 15	
16 to 20	
More than 20	

Frequency Graph

1. Using the graph function (“Insert” and “Column graph” for PC), choose the type of graph that best fits the type of data you are analyzing (most often a column graph).
2. (For Mac) Follow the prompts. You may manually select the information for the graph using the arrow and selecting from the Excel spreadsheet.
3. Make sure you label the axes and title appropriately.
4. Graph can be copied and pasted into a Word Document.

Median

1. For answers to the survey questions such as number of years of professional experience, designate a number for each criteria. Make a table, shown below, as you did for the frequency so that there is a record of this designation.
2. After selecting a vacant cell on the spreadsheet where you would like the results to be displayed, go to Function and select MEDIAN.
3. Select up to 255 numbers at a time either by manually selecting them in the spreadsheet (using the arrow at the right of the entry box) or by typing the cells (i.e. A1:A255). You will need to move to the next entry box (Number 2, Number 3, etc.) to enter more than 255 numbers. For example: A1:A30, A31:A61, A62:A92, etc.

Criteria (Yrs. Prof. Exp.)	Number
Less than 1	1
1 to 5	2
6 to 10	3
11 to 15	4
16 to 20	5
More than 20	6

4. Your results will be displayed as a number, and that number

corresponds with an answer (i.e. less than 1 year of experience is number 1). Make sure your final results indicate the criteria that the number stands for.

Mean

Your final calculation of the mean should be a number that falls within the data set. In the example above, the mean would fall between 1 and 6 depending on how many people gave responses for each answer. You should have a column that is (frequency X number). Then at the bottom of this column you would Sum all the values in this column and divide by N (total number of responses) and this will give you the Mean.

Mode

1. You are going to be using the same numbered list from the median calculations.
2. Click on the cell that you would like the results of the data to display in, go to function, and choose MODE.

Standard Deviation

1. (For Mac) Under "Tools," choose Data Analysis. Choose descriptive statistics.
2. (For Mac) When selecting the input data, choose rows that are numerical.
3. Remember that these results are based on the numbers that you assigned to criteria. Reporting the numbers alone may not indicate anything of importance about the data.

Part 3 – Detailed Statistics

Installing Data Analysis Toolpak on Excel

To install the Data Analysis Toolpak in Excel, go to the Microsoft Office button in the upper left corner of Excel; choose “Excel Options” as listed on the bottom of the menu. From the lefthand menu, choose “Add-ins.” At the bottom of this screen, select to change the add-ins. You are going to choose the Analysis Toolpak from the list.

This will allow you to do ANOVA, T-test, correlation coefficient, and Chi test.

When doing scatter plots make sure you use two variables (each with a number range that corresponds to a descriptive feature), and not frequency. They also should both have the same N. You will also need the Mean to be correctly calculated in order for the following data analyses to be correct.

Line Graphs- Scatter plot – tutorial link below

<http://www.brighthub.com/computing/windows-platform/articles/17778.aspx>

How to Make a Scatter Plot

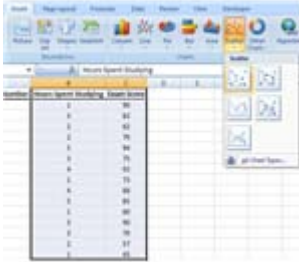
Step 1: Enter or copy/paste your data into an Excel [worksheet](#). See the instructions below that talk about the data in the Compassion Survey. As an example in this tutorial, we’ll be using data consisting of hours spent studying and final exam scores for a select group of students. (Click any image for a larger view.)

	A	B	C
1	Student ID Number	Hours Spent Studying	Exam Score
2	A1	1	90
3	A2	3	82
4	A3	2	62
5	A4	2	70
6	A5	5	94
7	A6	1	75
8	A7	4	92
9	A8	1	73
10	A9	4	88
11	A10	5	85
12	A11	1	90
13	A12	3	90
14	A13	1	70
15	A14	2	57
16	A15	1	45

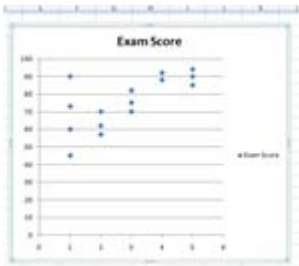
Step 2: Highlight the columns that contain the data you want to represent in the scatter plot. In this example, those columns are **Hours Spent Studying** and **Exam Score**.

	A	B	C
1	Student ID Number	Hours Spent Studying	Exam Score
2	A1	1	90
3	A2	3	82
4	A3	2	62
5	A4	2	70
6	A5	5	94
7	A6	1	75
8	A7	4	92
9	A8	1	73
10	A9	4	88
11	A10	5	85
12	A11	1	90
13	A12	3	90
14	A13	1	70
15	A14	2	57
16	A15	1	45

Step 3: Open the **Insert** tab on the Excel ribbon. Click on **Scatter** in the **Charts** section to expand the chart options box. Select the first item, **Scatter with only Markers**, from this box.



After making this selection, the initial scatter plot will be created in the same worksheet. You can resize this chart window and drag it to any other part of the worksheet. If you want to [move the chart to a new worksheet, click here for instructions](#).



Step 4: Make any formatting or design changes you wish in the **Design, Layout,** and **Format** tabs located under **Chart Tools** on the Excel ribbon.

For this data research assignment you need to consider:

1. Convert the Likert scores to numbers.
2. Choose (highlight) the two converted number variables for your data (as shown in the table below). You need to designate an X and a Y axis.

Individual data response	Years Professional Exp.	(X) Convert to number variable	(Y) Second variable converted
#1	Less than 1	1	2
#2	1 to 5	2	5
#3	6 to 10	3	2
#4	11 to 15	4	6
#5	16 to 20	5	5
#5	More than 20	6	6

Select Scatter plot. The scatter plot will appear.

3. With the scatter plot selected, go to Chart Type (at the top).

4. Scroll and select the chart that displays a line with $f(x)$.
5. This will give you the R-squared value, used with linear regression and least squares fit.
6. You may need to play around with the scatter plot so that it is easy to read and appropriately displays your data.

Least Squares Fit- Linear Regression

1. Under Data Analysis, choose Regression.
2. Select, either manually or by typing, the data for which you would like to calculate the results.

You will need to have matching N values from the two data sets in order to complete this task.

Correlation Coefficient

1. Highlight the cell in your spreadsheet in which you would like the results to appear.
2. Under Data Analysis, select Correlation.
3. For array 1, choose the column of transformed scores for the first factor (i.e. age).
4. For array 2, choose the column of transformed scores for the second factor (i.e. professional experience).
5. Make sure to label what this correlation coefficient is for (i.e. the correlation coefficient for age and years of professional experience).

T-test

1. Under Data Analysis, select T-test: Two-sample Assuming Equal Variances or T-test: Two-sample Assuming Unequal Variances.
2. Your results will appear in a table. Output range allows for you to place the results in the spreadsheet with you designating their destination. The other options allow for you to open the results in a new spreadsheet or new workbook. Your choice.

Chi Square Test

1. Use Function CHITEST.
See the handout on this test.

Three Factor ANOVA

1. Under Data Analysis, choose ANOVA: Three-Factor With Replication or ANOVA: Three-Factor Without Replication.

Choose the data for each input