

OPEN the Guide for Macs → Follow the instructions for StatPlus package!

Advanced Excel

Please download the data set:
<http://goo.gl/xgWOLm> (case sensitive)

- Find your course: Find your course:
- Advanced Excel Data

EASE
Essential Academic Skills Enhancement

UNM
STEM GATEWAY

Welcome to the EASE workshop series, part of the STEM Gateway program.

If you haven't already, please download the dataset from the EASE website. Don't worry about opening it right now, I'll make sure you know when to do so.

As with the Basic Excel course, this workshop will just begin to introduce you to additional things you can accomplish in Excel. This will focus on a few basic statistical analysis using the Data Analysis plug-in. The images are from the newest version of Excel, 2013 (which you can get through IT for free), for PC's. Older versions, or those in Macs may look slightly different, but the overall concepts should still apply. I want to make sure we are clear that this is by no means meant to be an all inclusive class in Excel. At each step, there are many options that I won't have time to go over, so I encourage you to spend some time on your own playing with them.

If you are using a Mac today, know that is perfectly ok, but I may not be able to help you with all the issues you encounter, so please help each other. Make sure you download the Mac Guide from the website, too, so you have it to follow along.

Lastly, as a heads up, at the end of the presentation you'll have to complete an exit

survey in order for me to accept your assessment sheet.

Refresh your memory - Basic Excel

- ▶ <http://goo.gl/xgWOLm> - EASE for Bio 203L
- ▶ Topics included:
 - ▶ Graphing
 - ▶ Bar Graphs (with error bars)
 - ▶ Scatter Plots (with trend lines)
 - ▶ Functions
 - ▶ Supplemental info on:
 - ▶ Formatting
 - ▶ Freeze panes
 - ▶ Filter
 - ▶ Sort

If you don't remember, the Basic Excel workshop that you took in 203 is available through the EASE workshop page. We are not going to spend time going over topics covered in that workshop, but many of them are beneficial for you here in 204, so make sure you go back and look through that presentation and supplemental information.

Topics that were covered include graphing, as well as the Filter and Sort functions, which will really help you this semester when it comes to cleaning up your dirty data before you can move forward with any analysis for your homework's.

Resources

- ▶ The EASE website (<http://goo.gl/xgWOLm>) for:
 - ▶ Instructions on getting FREE MS Office
 - ▶ This presentation
 - ▶ Mac instructions (including StatPlus)
 - ▶ Introduction to Statistics (Khan Academy)
 - ▶ Supplemental material: CountIf and IfThen statements

Some additional resources that are available to you through the EASE website are:

Instructions on how you can get Microsoft Office Pro for free for being a UNM student.

Just like with Basic Excel, this presentation is available as a PDF on the EASE website if you need to refer back to it. It includes the script, so you don't have to memorize everything I say to be able to interpret the slides.

There is also a file that guides you through the differences between PC and Mac versions of what we are about to go through, including the need for StatPlus, rather than the Data Analysis Add-In that we'll use today.

Lastly, there is a document with links to videos on basic Statistic concepts, if you want to learn more about what is underlying each of the tests we learned how to perform, as well as CountIf and IfThen statements.

Overview

- ▶ Vocabulary
- ▶ Analysis ToolPak add-in
- ▶ Statistics in Excel: Probability distributions
 - ▶ Regression, correlation
 - ▶ T-test
 - ▶ ANOVA

This workshop will first cover a few basic vocabulary terms that are important to understand when doing statistical analysis. I'll walk you through how to add and activate the Analysis ToolPak plug-in. Then we'll go over HOW to run three different statistical analysis Excel, but only briefly go into how to interpret the results.

Vocabulary: Mean, Median and Mode

Example Data: 13, 18, 13, 14, 13, 16, 14, 21, 13

- ▶ Mean (\bar{x}) = average: `=AVERAGE(CellFirst:CellLast)`
 $(13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13) \div 9 = 15$
- ▶ Median (\tilde{x}): rewrite the list in order
`=MEDIAN(CellFirst:CellLast)`
13, 13, 13, 13, 14, 14, 16, 18, 21
- ▶ Mode (Mo): `=MODE(CellFirst:CellLast)`
13

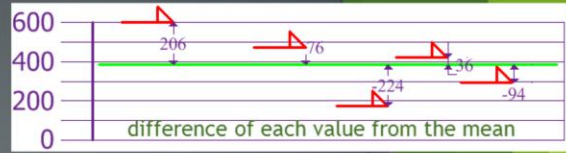
Let's go over mean, median and mode. If we have this example data set.

(*) To get the mean in Excel we type in `=average(first cell of the data: last cell of the data)`. The actual calculation is adding each data value together, then dividing by the number of data points.

(*) To calculate the median by hand, you have to re-write the data set in numerical order to find the central value. Since there are 9 values, the 5th value is the central point, and is 14 in this case. Again, I show you the Excel equation.

(*) Lastly, to calculate mode, it is the value to occurs the most in the data set. You can see from the Median example that we have 13 appear 4 times, 14 twice, and the other values only once. That means that 13 is the mode value.

Vocabulary: Variability



- ▶ Range: largest value - smallest value
`=MAX(CellFirst:CellLast)-MIN(CellFirst:CellLast)`
- ▶ Variance (σ^2):
average of the squared differences of each value from the mean
`=VAR.S(CellFirst:CellLast)`
- ▶ Standard Deviation (σ): measure of how spread out #s are;
the square root of Variance
`=STDEV.S(CellFirst:CellLast)`

Within the concept of Variability, the range is breadth of the data, that is, the largest value minus the smallest value. The equation in orange is what you would type into an Excel cell to calculate the range of a data set. The Maximum value within data in cell x through cell y, minus the minimum value within the data set.

(*) The variance is calculated as the squared difference of each value from the mean. So with this image, we would take the difference of each value (red lines) from the mean (green line) and square it, then divide it by the number of measurements, in this case 5. Again, the equation you'd enter into Excel is shown in orange. This is for variance of a sample set, rather than a population. We don't have time to go into what that means, but make sure you are using an appropriate calculation of variance within Excel, based on what is appropriate for your data set and test.

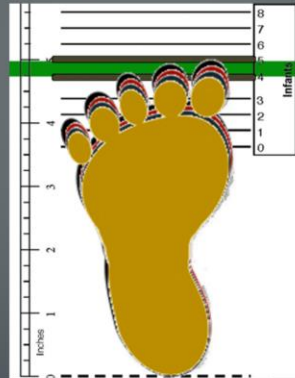
(*) Lastly, the Standard Deviation measures how spread out the values are from each other, it is calculated by obtaining the square root of the variance. As with variance, the standard deviation equation is for a sample set of values.

Types of Data

► Categorical: Qualitative/nominal

► Numerical: Quantitative

- Discrete
- Continuous



Before you can run any type of statistical analysis on your data, you have to first know what type of data you have. There are two major types:

(*) Categorical variables are those that are qualitative, or nominal. In other words, they do not consist of numbers directly, but of categories or names. For example, if we were comparing brands of shoes, the brand name is the categorical variable.

(*) Other attributes associated with shoes could be considered numerical, or quantitative, variables. These are measured on a numerical scale. These are split into two types: discrete and continuous.

Discrete numerical variables consist of only certain fixed values with no intermediate values, partial numbers are not possible, such as shoe size, you can have a 4 or a 5, but not a 4.27.

(*) On the other hand, continuous numerical variables include all possible intermediate values. Variables such as foot size, since your feet can have any value within a given shoe size.

Statistical Tests

- ▶ Determine if two data sets are significantly different or if they are correlated
 - ▶ Null hypothesis (H_0)
 - ▶ Alternative hypothesis (H_A)
- ▶ Probability Distributions to determine significance

Now that you know the type of data you have, you need to know what question you're asking. The statistical tests that we'll go through focus on determining if two or more data sets are significantly different from each other, or if the values are correlated.

The null hypothesis is that there no difference between data sets, or that the data sets are not correlated, that is, they are independent of each other. The alternative hypothesis accounts for all other possible outcomes.

To test the data in light of these hypotheses, we use a probability distribution table, which links each outcome of a statistical experiment with its probability of occurrence.

Determine Significance

- ▶ Level of significance (α) or confidence level (%)
- ▶ Test result: p-value
- ▶ Significant difference if $\alpha \geq p\text{-value}$

To be able to determine significance, we have to settle on a level of significance. How strong does the relationship have to be to reject the null hypothesis? The alpha is usually 0.01 (1%), 0.05 (5%), or 0.1 (10%). The smaller the value, the stronger the evidence must be to reject the null hypothesis. There are different levels because different disciplines have different requirements.

The probability that the outcome is not statistically supported assuming the null hypothesis is true is known as the **p-value**. P-values that are smaller than your set significance level (α) mean that the outcome is statistically significant. For example, a p-value less than 0.05 means there is less than a 5% chance that the observed outcome is due to chance.

If your p-value is less than the designated alpha, then there IS a statistical difference between the groups in question, at that level of confidence. In other words, you would reject the Null hypothesis, and accept the alternate hypothesis. We have an alpha to determine how acceptable in our results. The significance of the p-value is based on the chosen alpha. Yes, the p-value has to be small, but it is all relative to the chosen alpha. This is why medical fields use such a small alpha, but ecology allows a larger one.

Four Common Tests:

- ▶ Chi-Square: Comparing two categorical variables.
- ▶ Regression (Correlation): Comparing two non-categorical variables within the same data set.
- ▶ ANOVA and t-test: Comparing means and distributions across data sets. Comparing one categorical to one numerical variable.

There are four common tests to test these hypotheses:

Chi-Square tests compare two categorical variables. That's as much as we're going to discuss Chi-Square because there is not a "quick" way to use Excel to run this, like there is for the other types of tests we'll talk about.

Regressions compare two non-categorical variables within the same data set to determine if they are correlated or independent of each other.

And, ANOVA (ANalysis Of VAriance) and t-test compare means and distributions across two (t-test) or more (ANOVA) data sets. Comparing one categorical variable to one numerical variable.

Statistics: ANOVA vs. *t*-test

- ▶ They both require one categorical variable and one non-categorical variable.
- ▶ *t*-test compares averages (means) and distributions between two groups
- ▶ ANOVA compares means and distributions across more than two groups

As I mentioned, when you have Categorical and Numerical data, you have to select between an ANOVA and a *t*-test. These are used to compare the MEANS and distributions of data sets. If you want to compare medians, there are other tests for that.

The main difference between the two tests is that a *t*-test is used to compare the difference of means and distributions between TWO groups, where an ANOVA is used to compare the means of more than two groups. We'll go into details on each of these in a minute, but first let's look at some examples.

Selecting the correct analysis

		Variable 2	
		Categorical	Numerical
Variable 1	Categorical	Chi Square	<i>t</i> -test/ANOVA
	Numerical	<i>t</i> -test/ANOVA	Regression

Earlier, we went over how to determine the type of data you have, but not how that relates to which analysis you should run. This graphic should help with that choice. We'll go over how you can do each of these in Excel. You'll have to spend time looking through other resources I mentioned earlier to fully understand the mechanics behind each of these tests.

Match the Hypothesis with the Test

★ Assessment

Hypothesis	Variable 2	
	Variable 1	
Men and women are not, on average, the same height.	Categorical	Numerical
People from New Mexico, Texas and Nevada are not, on average, the same height.	Categorical	Chi Square
The temperature of the earth decreases with latitude.	Numerical	t -test/ANOVA
		Regression

So, based on what we've gone over so far, match the hypothesis with the appropriate test and write these down on your assessment sheet, with a description as to WHY you chose that answer.

What did you come up with? What type of test would you run for the first hypothesis?

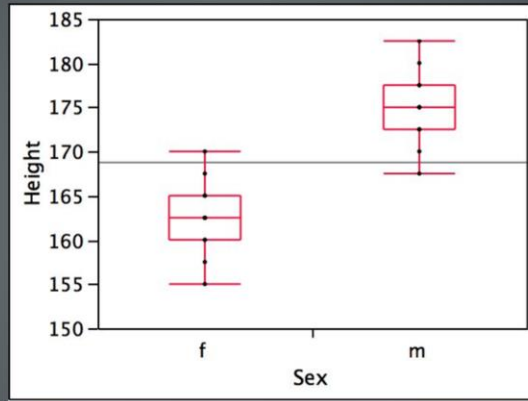
Match the Hypothesis with the Test

Hypothesis	Variable 1	Variable 2	Test
Men and women are not, on average, the same height.	Categorical	Numerical	t -test/ANOVA
People from New Mexico, Texas and Nevada are not, on average, the same height.	Categorical	Categorical	Chi Square
The temperature of the earth decreases with latitude.	Numerical	Numerical	Regression

The first hypothesis would need a t -test, since you are comparing categorical (gender) and numerical (height) variables. And, it is a t -test since there are only two groups being tested.

t-test results

Difference: 12.5cm
P<0.0001



When you run a t-test, you'll get a resulting p-value. Remember, the alpha level you've decided on will determine if the p-value indicates the acceptance or rejection of the null hypothesis. In this case the p-value is so small that the null is rejected at any of the standard alpha levels. That means there IS a difference between mean and distributions of male and female height.

This graphic is referred to as a box plot and the whiskers or error bars on the box are the upper and lower quartiles, or 1/4 and 3/4 medians. The error bars could also indicate the standard deviation, or range.

Just as a side note, you can NOT easily make box-plots like this in Excel, but I wanted you to see it in case you encounter it somewhere else, and to give you a visual of what the test is doing.

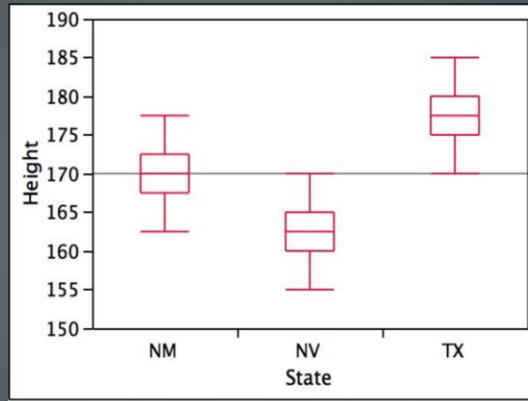
Match the Hypothesis with the Test

Hypothesis	Variable 2	
	Categorical	Numerical
Men and women are not, on average, the same height.		
People from New Mexico, Texas and Nevada are not, on average, the same height.	Categorical	t-test / ANOVA
The temperature of the earth decreases with latitude.	Numerical	t-test / ANOVA

For the second hypothesis, again we have categories of state, and numbers of height, but we have three groups, thus an ANOVA, rather than a t-test, is appropriate.

ANOVA results

Difference: 15cm
 $P < 0.0001$



As before, when you run an ANOVA, you'll get a resulting p-value. If the p-value is significant, it indicates that you reject the null that all means and distributions are the same, and that there IS a difference among the means and distributions of heights, but from this you can NOT determine WHICH pair are different, just that some pair is different. You'd have to run additional statistical tests to determine which pairs are different.

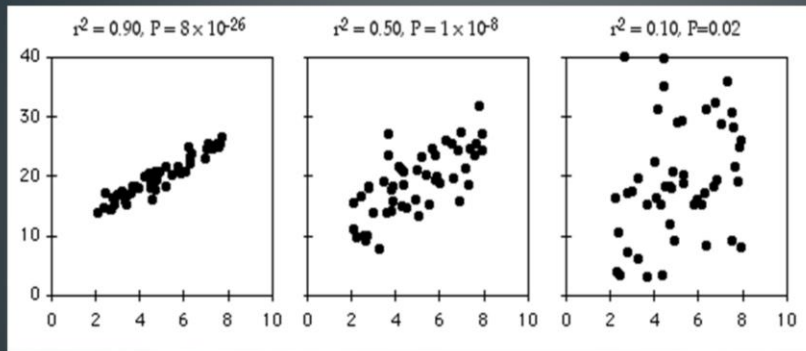
From looking at the graphic, it looks like NM and NV may be the same, as is NM and TX, but that NV and TX are different. But again, to statistically determine this, though, you'd have to run additional tests.

Match the Hypothesis with the Test

Hypothesis	Variable 1	Variable 2	Test
Men and women are not, on average, the same height.		Categorical	Numerical
People from New Mexico, Texas and Nevada are not, on average, the same height.	Categorical		Chi Square
The temperature of the earth decreases with latitude.	Numerical		t -test/ANOVA
			Regression

For the final hypothesis, since both temperature and latitude are numerical, a Regression is the best type of analysis.

Regression results



As with the other tests, you'll get a p-value, but there are other things to look at as well. The R-squared value indicates how well the line explains the data. A 0 indicates a poor explanation and a 1 is a perfect explanation. You can have a straight horizontal line that has a slope of zero, thus a high R-squared, but the p-value will be low. This is because the p-value indicates if the relationship of the two variables is different than 0, different than a straight vertical or horizontal line.

You can also get the regression equation, which is the equation of the trendline.

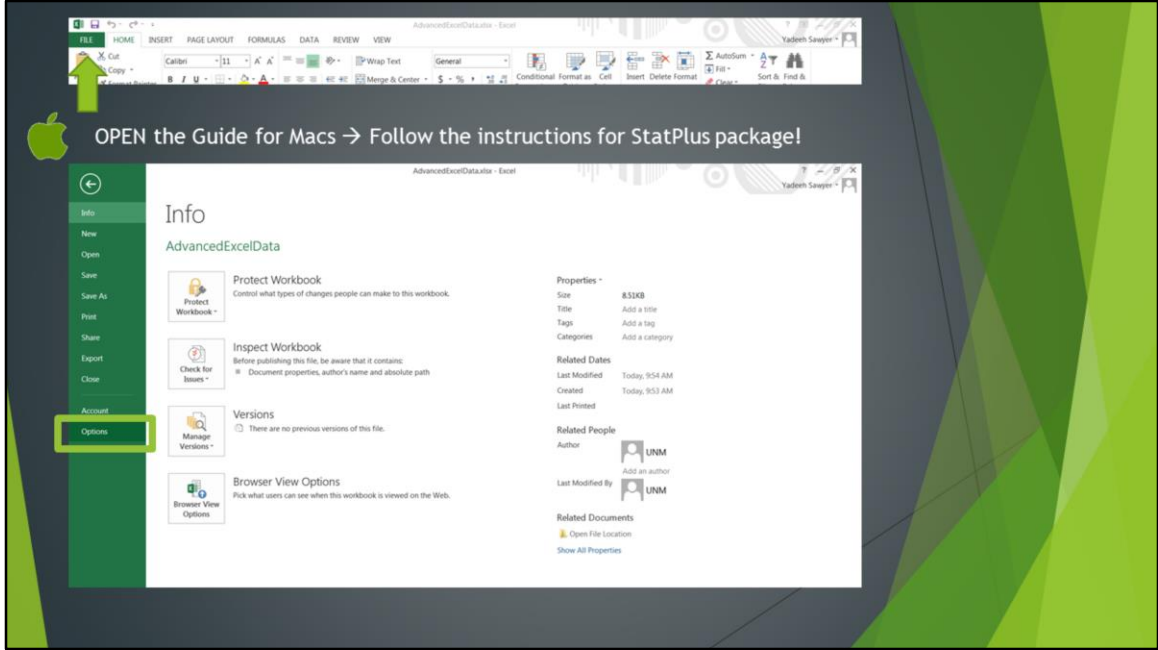
For this specific example, you can see the difference of a high, middle, and low R-squared values, with all significant p-values.

Analysis ToolPak add-in

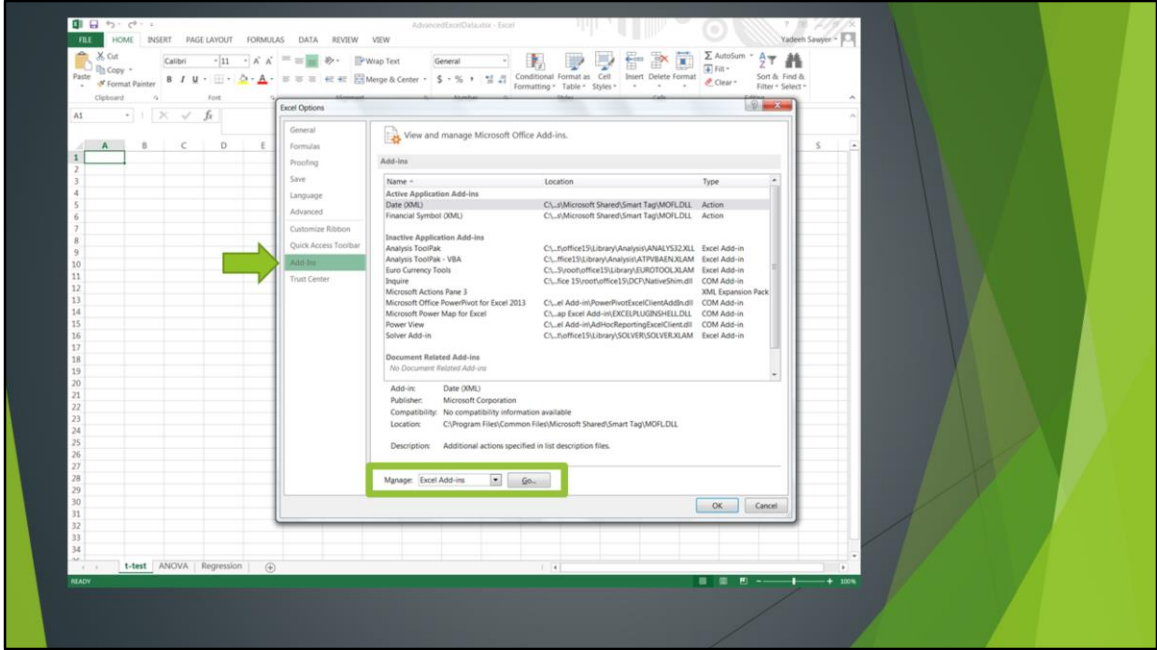
- ▶ Provides data analysis tools for financial, statistical, and engineering data analysis.

Ok, now let's get to you working with some data! You can run all of these tests within Excel without the plug-in by using the pull-down menu of functions under the "formulas" tab, but the ToolPak makes it easier to find and run.

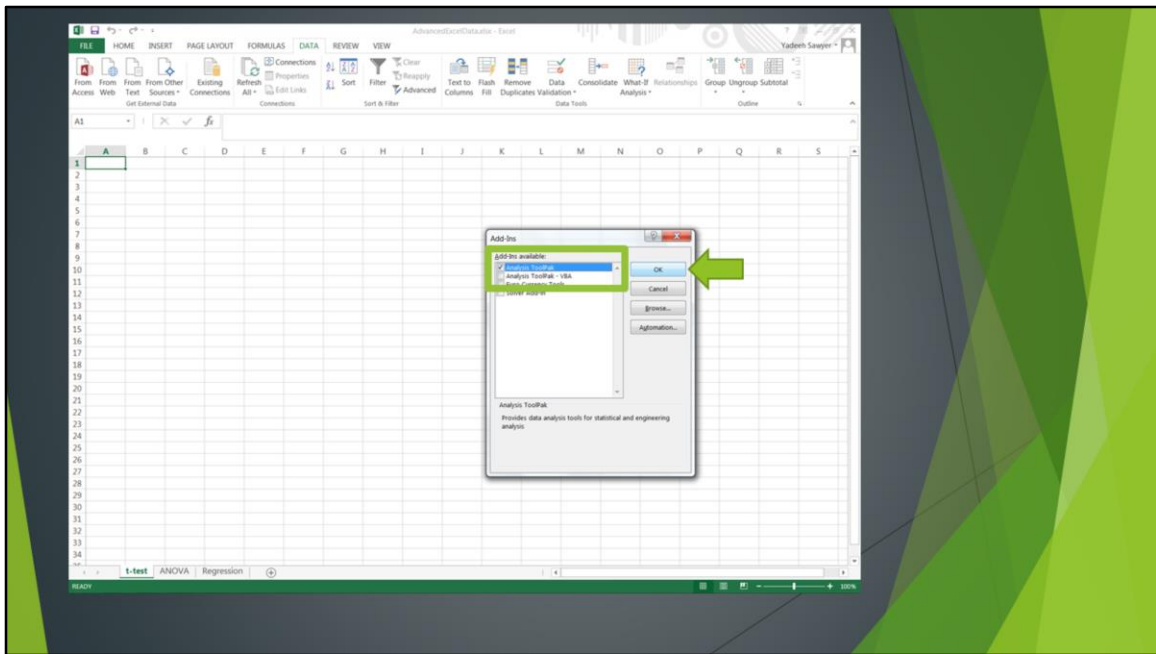
To more easily do any of these analysis in Excel, you have to activate the analysis toolpack. This provides data analysis tools for financial, statistical, and engineering data analysis. There are many more analysis than we are going to cover, so make sure you take some of your own time and look through them.



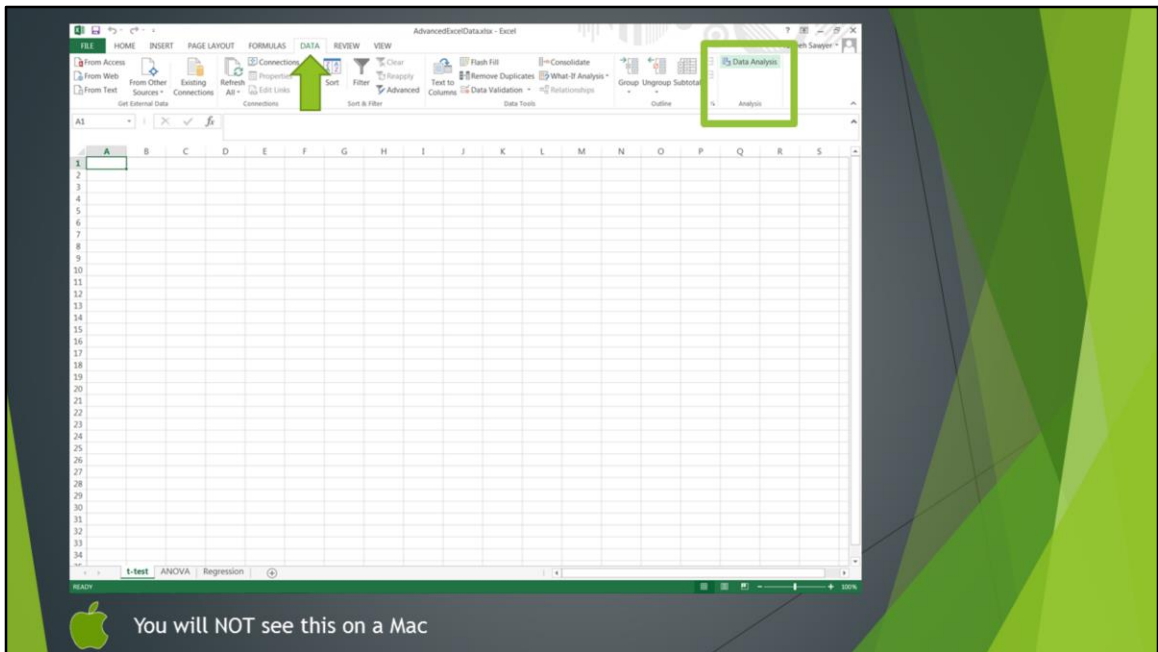
Ok, let's make sure the add-in is activated. Click "file", then select "options". Macs are slightly different, and you have to use StatsPlus, so make sure you go back and access the "Advanced Excel for Macs - Guide" to help you with the differences between platforms.



Within the options pop-out, select “add-ins”. You can see a list of activated add-ins, as well as the inactive ones. If the Analysis toolpak is not active, make sure “Excel add-ins” is selected under the manage field and click “go”.



You'll then see a list of add-ins that you can activate. Check the box next to the Analysis ToolPak and hit OK.



When you click the “data” tab, you should now see the “Data Analysis” option.

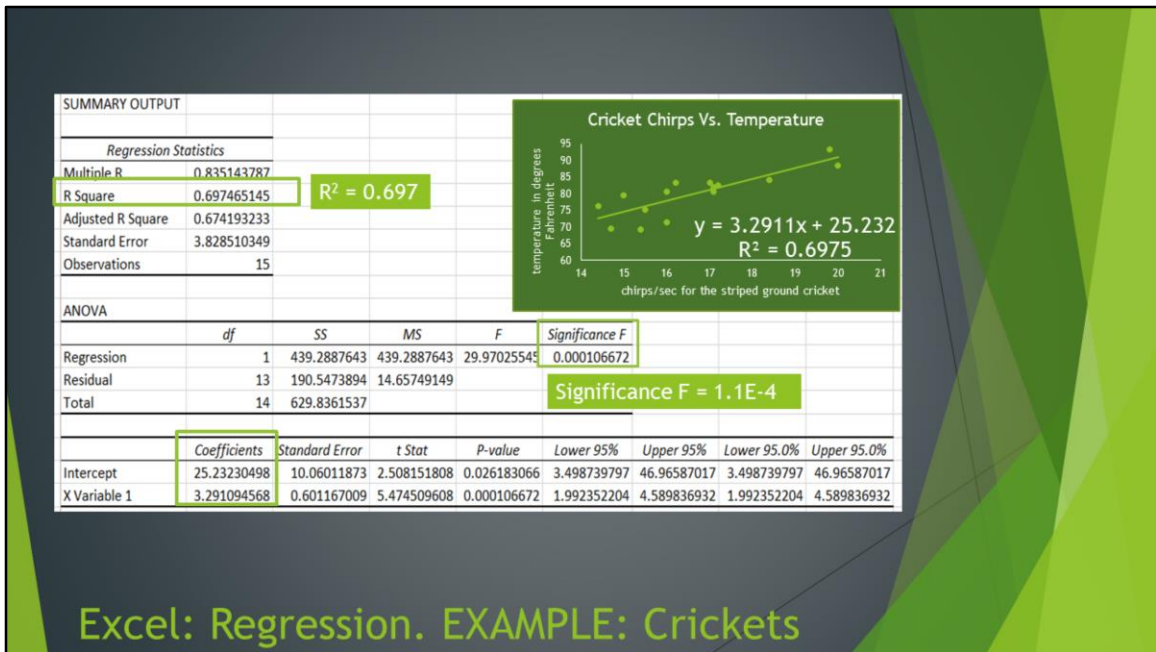
Once you’ve activated this on your personal computers, you should be set. But, if you are using a university computer, then you may have to repeat this process each time you use it. You can always check before you get in to your analysis by looking for the Add-In under the data tab.

Analysis within this are many, including descriptive statistics which will give you the Standard Deviation of a data set. Play around with them on your own at a later time, since, as I said already, we’ll only go over 3 major analysis.

Statistics: Linear Regression/correlation

- ▶ Looks at how strongly two variables are related (correlated). When there is a linear relationship between two variables, data points cluster more or less around a straight line.
 - ▶ H_0 : There is no relationship between two variables
 - ▶ H_A : There is a relationship between variables (either positive or negative)

The first type of analysis we're going to cover is a linear regression, or correlation. This looks at how strongly two variables are correlated or associated with each other. If the variables are correlated, they form a linear relationship. The Null hypothesis is that there is NO relationship between the two variables, that they are independent of each other. The alternate hypothesis then, is that there IS a relationship between the two variables.



Excel: Regression. EXAMPLE: Crickets

This Excel output example is for the frequency of cricket chirps in relation to the ambient temperature. The graph is not made automatically, but I wanted you to have a visual of the data in relation to the output statistics. And, step by step on how to create a graph like this was covered in the Basic Excel workshop.

The R-squared is 0.697. Remember, the closer to 1, the better the line explains the data.

For the regression equation, you look at the coefficients of the intercept and the x-variable, with the intercept being the final value, and x variable being the coefficient to x in your standard $y=mx+b$ equation of a line.

To determine if there is a significant relationship between the two variables tested, you look at the "Significance F" value, which assessed the same way you would a P-value. In this case, the results are significant, so you reject the null of no difference and accept the alternative hypothesis that there IS a correlation between cricket chirp rate and temperature. On a side note, remember, just because two variables are correlated, it does not mean that one is causing the other.

Open StatPlus with your excel sheet in view. Click on the Statistics Tab, then choose Regression, under regression, choose Multiple Linear Regression.

Excel: Regression

House #	Y: Actual Selling Price (\$1,000s)	X: House Size (100s ft ²)
1	89.5	20
2	79.9	14.8
3	83.1	20.5
4	56.9	12.5
5	66.6	18
6	82.5	14.3
7	126.3	27.5
8	79.3	16.5
9	119.9	24.3
10	87.6	20.2
11	112.6	22
12	120.8	19
13	78.5	12.3
14	74.3	14
15	74.8	16.7

Ok, now it's your turn. Make sure you are on the Regression worksheet. Select the data tab. Then click on "Data Analysis".

One thing I want to point out is the "HELP" option, which is great if you want a description of each test to determine if it's appropriate.

Scroll down and select the Regression analysis.

These data are for the relationship between the selling price of a house and the size of the house.

For variables, choose the grid icon. Once you highlight what you want, click back on the StatPlus popup menu

For the confidence interval, click preferences in the popup menu

- Y and X input range
- Labels
- Confidence level = 95% (=α of 0.05)
- Output worksheet name

★ Assessment

Excel: Regression

Before you can run your regression, you'll need to determine which variable is the independent (x-axis) and which is the dependent (y-axis) variable.

Now, as you did when you selected the desired data in the Basic Excel workshop, you select the desired data for the input Y and X ranges. Make sure "labels" is selected if you want to include the headers, which is convenient for output labeling.

Set the confidence, or alpha level to the level you want, and enter the name you want for the output worksheet, then hit OK.

(*) Answer these questions on your assessment: What is the R-squared? Is the regression significant? What does this tell you about the relationship between house selling price and size?

Statistics: *t*-test

- ▶ Compares means and distribution of two data sets or groups
 - ▶ Many types, focus two-tail *t*-test
 - ▶ H_0 : The sample mean or distribution does not differ from the population mean
 - ▶ H_A : The sample means or distribution is different from the population mean

Next, let's start with a *t*-test first. As I said, this compares two means and distributions. As with many of the topics we're covering, there are various types of *t*-tests, but we're going to focus on two-tail *t*-tests.

Again, the null hypothesis is that there is no difference between the means and distributions, while the alternate hypothesis is that there IS a difference between the means or distributions.

t-Test: Two-Sample Assuming Unequal Variances		
	Older Adults	Younger Adults
Mean	44.5	28.1
Variance	75.38888889	72.98888889
Observations	10	10
Hypothesized Mean Difference	0	
df	18	
t Stat	4.257546666	
P(T<=t) one-tail	0.000236832	
t Critical one-tail	1.734063607	
P(T<=t) two-tail	0.000473663	
t Critical two-tail	2.10092204	

Excel: *t*-test. EXAMPLE: Life Satisfaction

When you do a *t*-test in Excel, this is what the output will look like. You can see the mean and variance, as well as the number of data points. The default hypothesis is that the mean difference is zero, but you do have the ability to change this. You can see the degrees of freedom listed, as well as various test statistics. We're focused on the two-tail *p*-value, which in this case is 0.000474. Assuming our chosen alpha is 0.05, the null hypothesis is rejected because there is a significant difference between life satisfaction in older adults versus younger adults.

t-Test: Two sample assuming unequal variance

Choose Basic Statistics, then choose Compare Means (T-test) → Change the t-test type to “t-test assuming unequal variances” from the dropdown menu at the bottom of the pop up window.

Excel: t-test

Group	Scores
1	20
1	23
1	32
1	24
1	25
1	28
1	27.5
2	25
2	46
2	56
2	45
2	46
2	51
2	34
2	47.5

Make sure you are on the t-test worksheet. This is data for two different groups and their exam scores.

Select the data tab. Then click on “Data Analysis”.

Scroll down to the t-test and, for the simplicity of this workshop, select the two sample assuming unequal variance. This is the conservative approach, versus assuming equal variance.

Click OK.

Group	Scores
1	20
1	23
1	32
1	24
1	25
1	28
1	22.5
2	25
2	46
2	56
2	45
2	46
2	51
2	34
2	47.5

- Value range
- α level = 0.05
- Output worksheet name

★ Assessment

Excel: t-test

You'll have to select the numerical (not categorical) data for each data set, in this case you are looking group 1 and group 2 scores. If you leave the hypothesized difference blank, it will assume the null hypothesis is no difference.

As before, if you have header rows, that's when you'd select tables, and make sure your alpha level is at the level you want it to be.

Lastly, enter a name for the output worksheet, something like t-test results.

Hit OK.

(*) A new worksheet is created with the results. Take a second and answer the question on your assessment sheet. What is the resulting p-value of the t-test? And, based on this, are the means of the two groups significantly different from each other?

Statistics: ANOVA (Analysis of Variance)

- ▶ Used for comparing the means and distributions of more than two groups
 - ▶ H_0 : The means or distributions of all groups are all equal
 - ▶ H_A : At least one of the means or distributions are different (or, not all equal)

Lastly, we'll go over ANOVA, which is essentially an extension of the t -test. It is used for comparing the means and distributions of more than two groups. Unlike the t -test, an ANOVA determines if all the means and distributions are equal or at least one of the means or distributions is different. The test does not tell us which one is different. You have to conduct additional tests to determine which pairs are different.

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
1	6	3320	553.3333333	12133.86667		
2	6	3416	569.3333333	2302.666667		
3	6	3663	610.5	3593.5		
4	6	2791	465.1666667	3318.566667		
5	6	3664	610.6666667	3455.466667		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	85356.46667	4	21339.11667	4.301535904	0.008751641	2.75871047
Within Groups	124020.3333	25	4960.813333			
Total	209376.8	29				

Excel: ANOVA. EXAMPLE: Concrete

As before, this is what the output will look like for an ANOVA. This is from a comparison of water retention in 5 different brands of concrete. You can see the mean and variance, as well as the number of data points for each group. The p-value between groups is 0.0088, so the null hypothesis is rejected and there IS a significant difference between at least one pair of values. However it doesn't tell you where the significant difference is.

The screenshot shows the Microsoft Excel interface with the Data Analysis ToolPak dialog box open. The dialog box is titled 'Data Analysis' and has 'Anova: Single Factor' selected in the 'Analysis Tool' list. The spreadsheet data is as follows:

Effect of Feed type on animal weight gain	A	B	C	D
3	60.8	78.7	92.6	86.9
4	67	77.7	84.1	82.2
5	54.6	76.3	90.5	83.7
6	61.7	79.8		90.3

Below the dialog box, there is a green box with the text 'Anova: Single Factor'. Below that, there is a text instruction: 'Choose Analysis of Variance (ANOVA). Choose 1 -way ANOVA.' At the bottom left of the Excel window, there is a small green Apple logo and a green box containing the word 'ANOVA'. At the bottom center, the text 'Excel: ANOVA' is displayed in a large green font.

Ok, your turn now. Make sure you are on the ANOVA worksheet. These data are for the effects of feed type on animal weight gain. Select the Anova: Single Factor test in the Data Analysis ToolPak.

Effect of Feed type on animal weight gain

	A	B	C	D
1	Effect of Feed type on animal weight gain			
2	A	B	C	D
3	60.8	78.7	92.6	86.9
4	67	77.7	84.1	82.2
5	54.6	76.3	90.5	83.7
6	61.7	79.8	90.3	

ANOVA: Single Factor

Input Range: \$A\$2:\$D\$7

Grouped By: Columns

Labels in First Row:

Alpha: 0.05

Output options:

Output Range:

New Worksheet In: ANOVA_output

New Workbook:

★ Assessment

Excel: ANOVA

The input range will be for the entire data set, and if you have headers (in this case, A-D for type of feed), make sure the correct alpha level is indicated, and make a name for the output worksheet. Hit OK.

(*) on your assessment sheet, what is the resulting p-value of the ANOVA? Are ANY of the groups significantly different? And, are ALL of the groups significantly different?

Resources

- ▶ The EASE website (<http://goo.gl/xgWOLm>) for:
 - ▶ Basic Excel
 - ▶ Instructions on getting FREE MS Office
 - ▶ This presentation
 - ▶ Mac instructions (including StatPlus)
 - ▶ Introduction to Statistics (Khan Academy)
 - ▶ Supplemental material: CountIf and IfThen statements

Some additional resources that are available to you through the EASE website are:

Instructions on how you can get Microsoft Office Pro for free for being a UNM student.

Just like with Basic Excel, this presentation is available as a PDF on the EASE website if you need to refer back to it. It includes the script, so you don't have to memorize everything I say to be able to interpret the slides.

There is also a file that guides you through the differences between PC and Mac versions of what we are about to go through, including the need for StatPlus, rather than the Data Analysis Ad-In that we'll use today.

Lastly, there is a document with links to videos on basic Statistic concepts, if you want to learn more about what is underlying each of the tests we learned how to perform, as well as CountIf and IfThen statements.

Refresh your memory - Basic Excel

- ▶ <http://goo.gl/xgWOLm> - EASE for Bio 203L
- ▶ Topics included:
 - ▶ Graphing
 - ▶ Bar Graphs (with error bars)
 - ▶ Scatter Plots (with trend lines)
 - ▶ Functions
 - ▶ Supplemental info on:
 - ▶ Formatting
 - ▶ Freeze panes
 - ▶ Filter
 - ▶ Sort

If you don't remember, the Basic Excel workshop that you took in 203 is available through the EASE workshop page. We are not going to spend time going over topics covered in that workshop, but many of them are beneficial for you here in 204, so make sure you go back and look through that presentation and supplemental information.

Topics that were covered include graphing, as well as the Filter and Sort functions, which will really help you this semester when it comes to cleaning up your dirty data before you can move forward with any analysis for your homework's.

Questions?

EASE Workshop Series
Essential Academic Skills Enhancement
<http://goo.gl/lcha0y>

UNM
STEM GATEWAY

Any questions?

For the survey, on your data set, there is a tab that says “Exit Survey” that will take you to it. When you are done with that and your assessment, please call me to see the confirmation page for the survey and I’ll take your assessment, then you are free to go.