# LAB NOTES #9

## I RECOMMEND NOT PRINTING THESE LAB NOTES!!

## They will be available throughout the course for your reference

Review and welcome back from Spring Break!

If measured to enough precision, there will always be a difference between two things/groups/samples. It may be 0.000001 grams or 124 pounds, but there will be a difference. The null hypothesis states there is no **statistical** difference, greater than the level of error you are willing to accept of making a Type I error. A Type I error is made when you state there is a difference and there really isn't one.

Alpha level is the risk you are willing to take of making a Type I error. By convention, it is almost always 5% (0.05)

With an alpha level set at 0.05, a p-value of 0.05 or less would lead you to reject the null hypothesis. That probability (known as the p-value and found in SPSS as *Sig.)* of finding, by chance, a difference as big as you found is less than your threshold level so you are confident that there truly is a difference. And you determined your risk tolerance (set your alpha level) **before** doing the data collection.

As an example, we have been looking at the K-S test as a way to evaluate the distribution of data for normality. This test compares your distribution to a theoretical normal distribution. The results of the two -tailed test are shown as the **last item** in your K-S results output box.

If there is no difference between the distributions (with an alpha level set at 0.05), then the p-value would have to be greater than 0.05. (i.e. you are saying that the probability is greater than the one you are willing to take that the difference is due to chance.) You "accept" (Dirty word, - don't use it) that yours is normal. What you actually are doing is "failing to reject the null" (Good words – use them.)

However, if the p-value is 0.05 or less (again, if you have set your alpha level at 0.05) then you would reject the null. You state that there is a difference between your distribution and a theoretical normal one.

The alpha level and p-value will be used throughout your comparisons today

Mean is a good choice in a normal distribution. In a skewed distribution, consider the median (middle number): half are above, half are below, and the extremes have a smaller influence.

From last week's data, what is the most frequent number of accidents in a five year period? Mode answers this. Why does the insurance company want to know? I would say that they want to recoup their costs by assessing premiums - so spreading the cost based on the mean accident rate would do this. However, they also don't want to lose customers, and if you see your rates go up after no accidents in 5 years (more than a 1/4 of the women if I remember correctly) you may be inclined to switch insurance companies. What you might want to do is charge the ones who have the most accidents the most money - a great reactive strategy. Would it be nice to predict who would have accidents? Regression analysis helps do this and we will practice later in the term.

Today, let's go straight into the homework, which is a bit longer, and a review.

**Homework 9**

Introduction

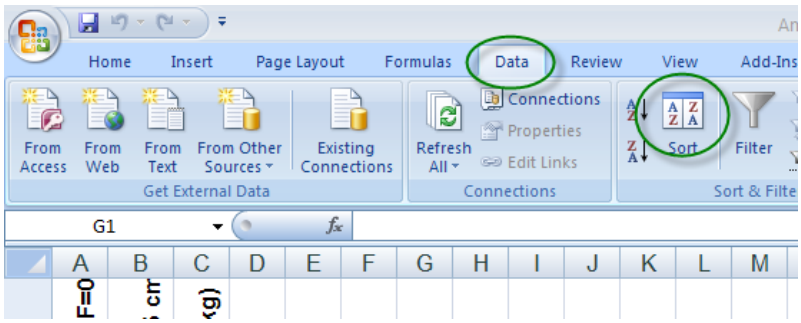
After the introduction, in this lab you will do a series of tasks. There are two parts of the lab, a Perform section and a Send section. Each Perform section task will end with something to do in the Send section and will then tell you where to go next.

Use the data found in the files *Anthropometrics* and *10K* on Blackboard. Open these files and save each file to desktop. Close Blackboard. SPSS and Blackboard have been known to not get along.

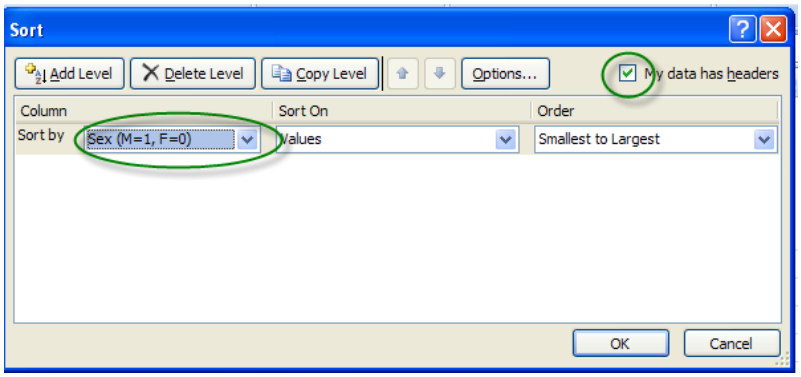
You will be able to perform the Homework efficiently if you sort the data before importing it into SPSS . Make a version 2 of the Excel file *Anthropometrics*, name it some way to differentiate it, then open this new version Excel file directly (not through SPSS ) and sort the data.

Here is what I suggest you do to the Excel file.

Open the file you just renamed. Click on the data tab at the top, then on Sort



Sort by Sex.




You have the data now all sorted so that the women are all together and the men are all together. Now cut the bottom 50 subjects from all three columns and paste these next to the first 50. You now have column 1 and 4 sex, column 2 and 5 height, and column 3 and 6 mass. Rename the columns so they are unique and descriptive. (men and women!)

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Women | WomenHeight | WomenMass | Men | MenHeight | MenMass | |
| 2 | 0 | 161.0 | 72.6 | 1 | 185 | 84.5 | |
| 3 | 0 | 167 | 77.5 | 1 | 185 | 92.1 | |
| 4 | 0 | 166 | 79.3 | 1 | 187 | 99.4 | |
| 30 | 0 | 166.5 | 66.1 | 1 | 184 | 114 | |
| 31 | 0 | 156 | 75.7 | 1 | 180 | 108 | |
| 32 | 0 | 160.5 | 55 | 1 | 181 | 89.8 | |
| 33 | 0 | 166 | 71.1 | 1 | 183 | 108 | |
| 34 | 0 | 168 | 76.7 | 1 | 182 | 101 | |
| 35 | 0 | 152.1 | 50.1 | 1 | 169 | 91 | |
| 36 | 0 | 162.5 | 55.1 | 1 | 171 | 93.1 | |
| 37 | 0 | 166.5 | 68 | 1 | 170 | 101 | |
| 38 | 0 | 158.5 | 55.3 | 1 | 179 | 92.5 | |
| 39 | 0 | 160.5 | 64.2 | 1 | 182 | 84.4 | |
| 40 | 0 | 181 | 73.3 | 1 | 174 | 91.8 | |
| 41 | 0 | 156 | 74 | 1 | 171 | 84.5 | |
| 42 | 0 | 163 | 84.4 | 1 | 182 | 93.3 | |
| 43 | 0 | 161.5 | 78.4 | 1 | 178 | 84.3 | |
| 44 | 0 | 154.5 | 55.5 | 1 | 183 | 117 | |
| 45 | 0 | 162 | 87.3 | 1 | 178 | 95.7 | |
| 46 | 0 | 163 | 66.4 | 1 | 181 | 92 | |
| 47 | 0 | 164 | 70.7 | 1 | 172 | 92.3 | |
| 48 | 0 | 157.5 | 58.5 | 1 | 161 | 73.5 | |
| 49 | 0 | 164.5 | 62.4 | 1 | 177 | 80.8 | |
| 50 | 0 | 160 | 66.2 | 1 | 185 | 95.6 | |
| 51 | 0 | 165.5 | 52.1 | 1 | 181 | 75.1 | |

Split scrren is on here just to show how my sorted data look.

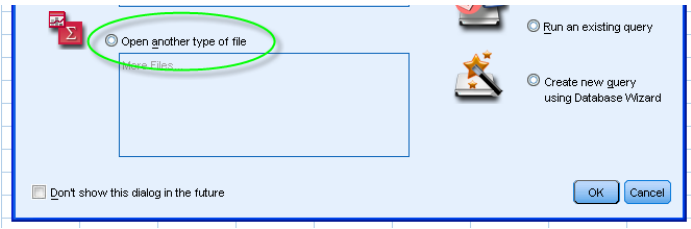
Save and Close this modified Excel file

Now we are ready for SPSS .

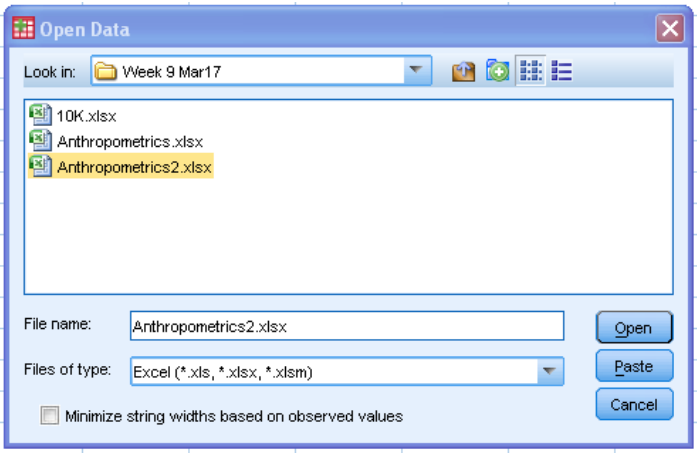
Open SPSS, *Open an existing data source. More files…* will be shaded grey in the box under the selection Open another type of file. (See green oval below). Select this.

Now we are ready for SPSS .

Open SPSS, Open an existing data source. More files… will be shaded grey in the box under the selection Open another type of file. (See green oval below). Select this.

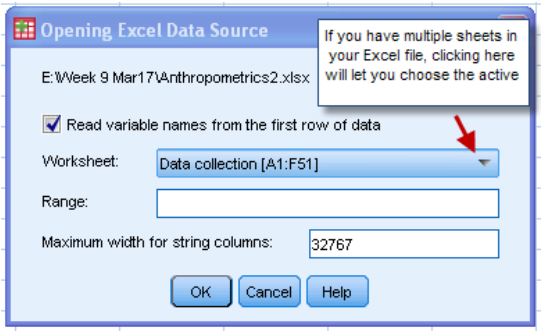



Click OK. Look in Desktop. Can't find it? In the dialog box, select the *Files of type* drop down box and click on *Excel*.

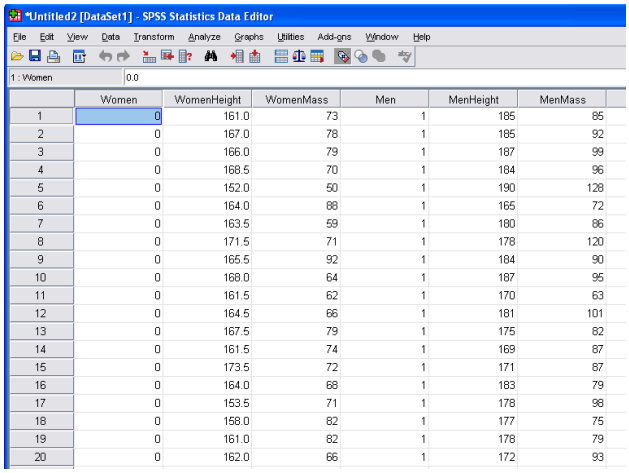



Now you should see your file (the one you just modified). Select the file. Click **Open**.

This will open another dialog box. *Read variable names from the first row of data* should be checked.




Click **OK**

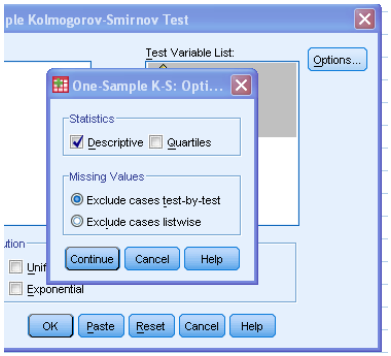



| | Women | WomenHeight | WomenMass | Men | MenHeight | MenMass |
|---|---|---|---|---|---|---|
| 1 | 0 | 161.0 | 73 | 1 | 185 | 85 |
| 2 | 0 | 167.0 | 78 | 1 | 185 | 92 |
| 3 | 0 | 166.0 | 79 | 1 | 187 | 99 |
| 4 | 0 | 168.5 | 70 | 1 | 184 | 96 |
| 5 | 0 | 152.0 | 50 | 1 | 190 | 128 |
| 6 | 0 | 164.0 | 88 | 1 | 165 | 72 |
| 7 | 0 | 163.5 | 59 | 1 | 180 | 86 |
| 8 | 0 | 171.5 | 71 | 1 | 178 | 120 |
| 9 | 0 | 165.5 | 92 | 1 | 184 | 90 |
| 10 | 0 | 168.0 | 64 | 1 | 187 | 95 |
| 11 | 0 | 161.5 | 62 | 1 | 170 | 63 |
| 12 | 0 | 164.5 | 66 | 1 | 181 | 101 |
| 13 | 0 | 167.5 | 79 | 1 | 175 | 82 |
| 14 | 0 | 161.5 | 74 | 1 | 169 | 87 |
| 15 | 0 | 173.5 | 72 | 1 | 171 | 87 |
| 16 | 0 | 164.0 | 68 | 1 | 183 | 79 |
| 17 | 0 | 153.5 | 71 | 1 | 178 | 98 |
| 18 | 0 | 158.0 | 82 | 1 | 177 | 75 |
| 19 | 0 | 161.0 | 82 | 1 | 178 | 79 |
| 20 | 0 | 162.0 | 66 | 1 | 172 | 93 |

(**If** you cut and paste from Excel, make sure the right level of precision is visible in Excel!)

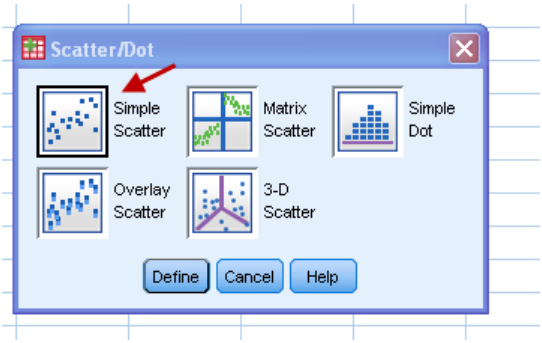
# Perform

Perform #1. Perform a one sample K-S test to see if these data are normal: height of men, the height of women, the mass of men, and the mass of women. In *Options* click on *Descriptive* under *Statistics.* (Hint: If you use the modified Excel file, this can be a ONE analyze event). Go to



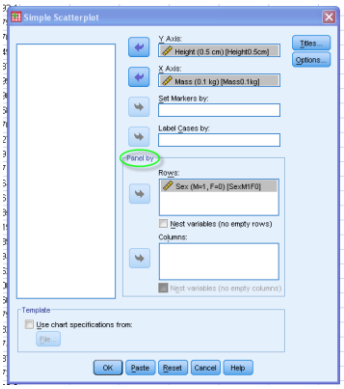

Go to Send #1.(look below to page 8)

Perform #2. Combine the data in your Excel spread sheet back into three columns: Sex, Height, Mass. (This was your original Excel download) Start SPSS again and import as before. Do a scatter plot [Graphs…, Legacy Dialogs… ], pick Simple Scatter.




Put height on the Y axis and mass on the X axis.

Panel by sex.

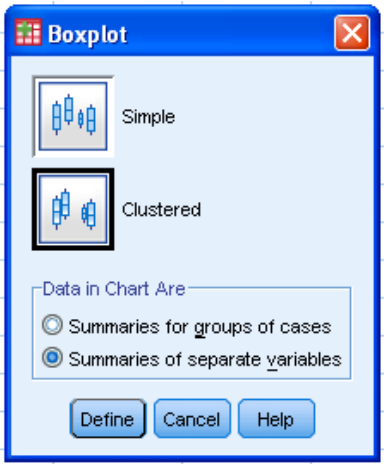


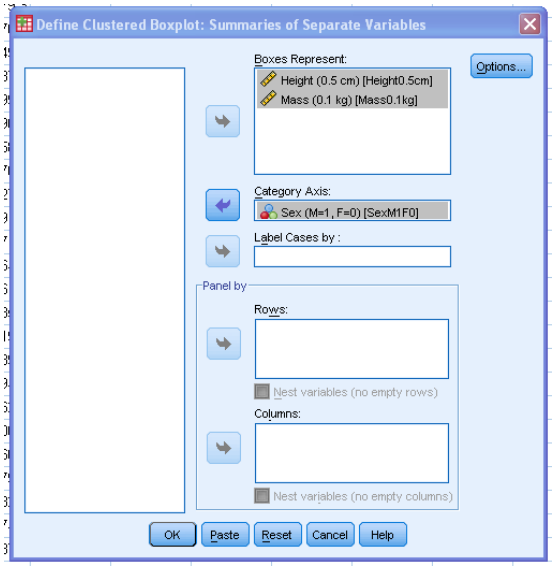
Do this twice: first by columns, then by rows (or vice versa).

Which is easier on the eyes?  No right answer here.

Now do a Box and Whisker Plot and put both height and mass in.

This will be a **Clustered** boxplot.

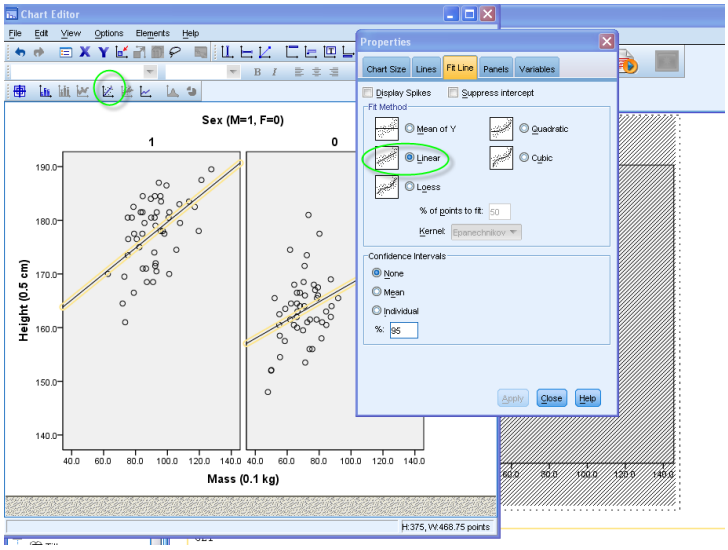





Which do you prefer for mass and height? Boxplot or scatter plot? Why? **Don't delete anything from SPSS.**

Go to Send #3.

Perform #3. Scatter has some nice relationship/association tools. Open a chart editor in the scatter plot and put in an association line (Called 'Add Fit Line at Total'). Click on Linear.




Go to Send #4.

Perform #4. 400 runners were randomly chosen and assigned to one of four training groups (four groups of 100) to improve their 10,000 meter time. The data in the *10K.xls* are the improvement times (to the 100th of seconds) for four training interventions, one for each of the four groups. Assume the data are normal. They are ratio scale, right?

Compare group means for improvement in 10K run time.

Don't forget the post hoc (Use Tukey)

Go to Send #5

# Send

Send 1.

a. The Descriptive table and **JUST** the Asmyp Sig (2 tailed) values from the output.

b. Tell me if they are or aren't normally distributed values and …

c. …how you know this.

Go to Send #2.

Send 2.

a. Tell me if you agree or don't agree that this measure of central tendency and these measures of variability are adequate?

b. Tell me **why** you agree or don't agree that this measure of central tendency and these measures of variability are adequate

c. If you don't think they are adequate, run the appropriate descriptive and send them as well.

Go to Perform #2

Send 3.

a. Copy the chart you prefer into your Word document and….

b. **….tell me what you like about it**. (Why it is a good choice to convey this information)

Go to Perform #3.

Send 4.

a. Copy and paste this into your Word document. Close SPSS.

b. Tell me in words what Association is when we are talking statistics

Go to Perform #4.

Send 5.

a. Paste the results of your analysis of the change scores.

b. Tell me in words which group(s), if any, improved significantly over any group and by how much.

c. Tell me in words what the p-value is compared to in order to decide whether statistical significance was reached when you compared the means

Send it to me in the usual format