**Stat 301 A Lab 2**

# Overview

Use commands in the JMP Analyze / Distribution menu to produce quantitative summaries of single samples.

Use Graph / Graph Builder to draw a box plot.

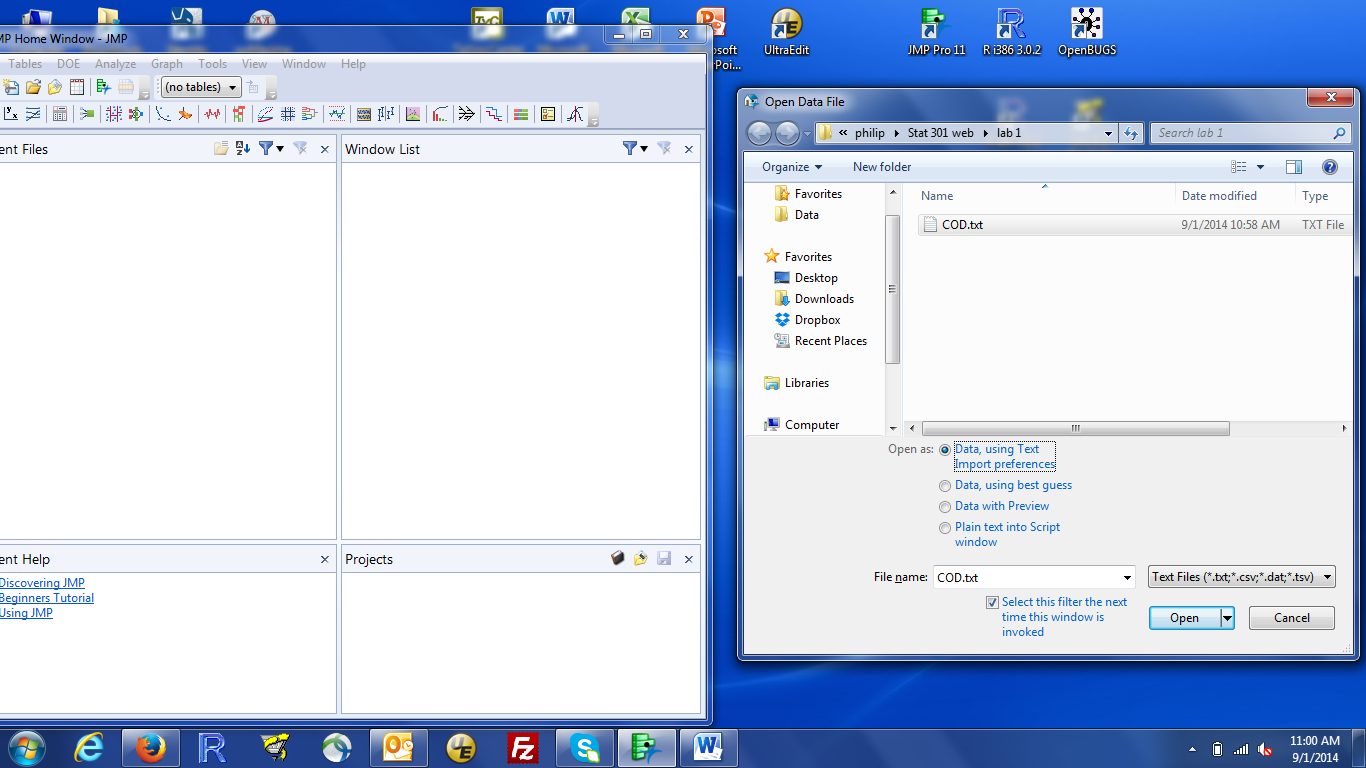
Learn how to save JMP graphs or other JMP output in a word processor.

Demonstrate the importance of the Modeling type attribute.

# Lab Activities

1. We will recreate the analysis of the COD data presented in class. That file is COD.txt on the class web site, <http://www.public.iastate.edu/~pdixon/stat301/>

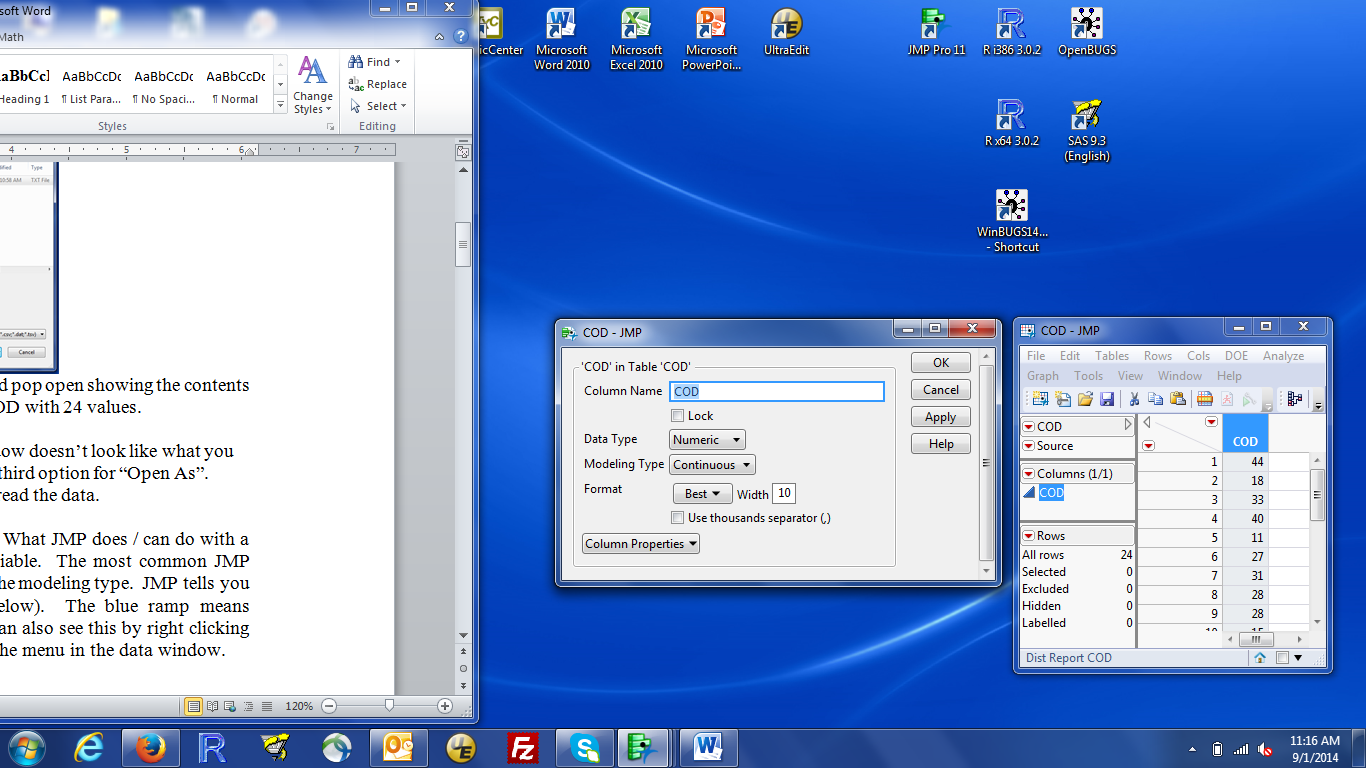
Download that file into a convenient folder and use File / Open (change the file type from “All JMP files” to Text files), select the file, and check “Data, using Text Import preferences”. The File / Open dialog window should look like:



then left click the open button. A new window should pop open showing the contents of the COD data set. This has one variable, called COD with 24 values.

If the data was not read properly (the data in the window doesn’t look like what you expected), reread the using “Data, with preview” the third option for “Open As”. This gives you the ability to tell JMP exactly how to read the data.

1. The “Modeling Type” of a column is critical to JMP. What JMP does / can do with a variable depends on the modeling type for the variable. The most common JMP problems happen when JMP and you disagree about the modeling type. JMP tells you the modeling type in the “Columns” box (see below). The blue ramp means quantitative data (continuous in JMP’s lingo). You can also see this by right clicking on the column, or selecting Cols/Column info from the menu in the data window.



There are two related boxes in the Cols/Column Info dialog box. “Data Type” is what is contained in the column (numbers, characters) and “Modeling Type” is how to use that information. Numbers can be quantitative = continuous or qualitative = nominal; characters can only be nominal (What is the average of A, B, C?).

Make sure COD was read as a continuous variable.

1. Make the JMP home window active, then choose Analyze / Distribution. Select COD and make it a Y variable by clicking “Y, columns”. Then click OK to do the analysis.
2. The results window (below) shows a histogram next to a “outlier” box plot, a table of quantiles, and a Summary Statistics box with 6 numbers:

Mean = sample average,

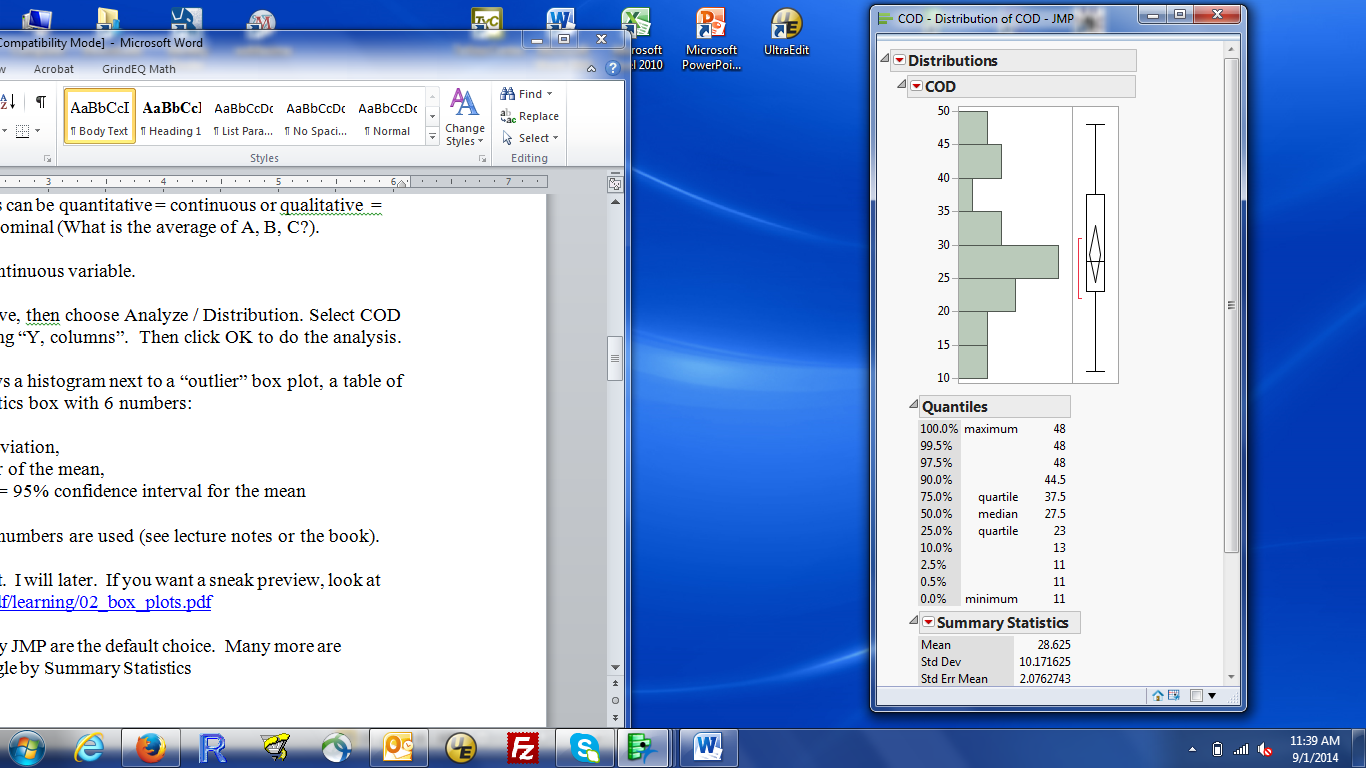
Std Dev = sample standard deviation,

Std Err Mean = standard error of the mean,

Upper and Lower 95% Mean = 95% confidence interval for the mean

N = number of observations

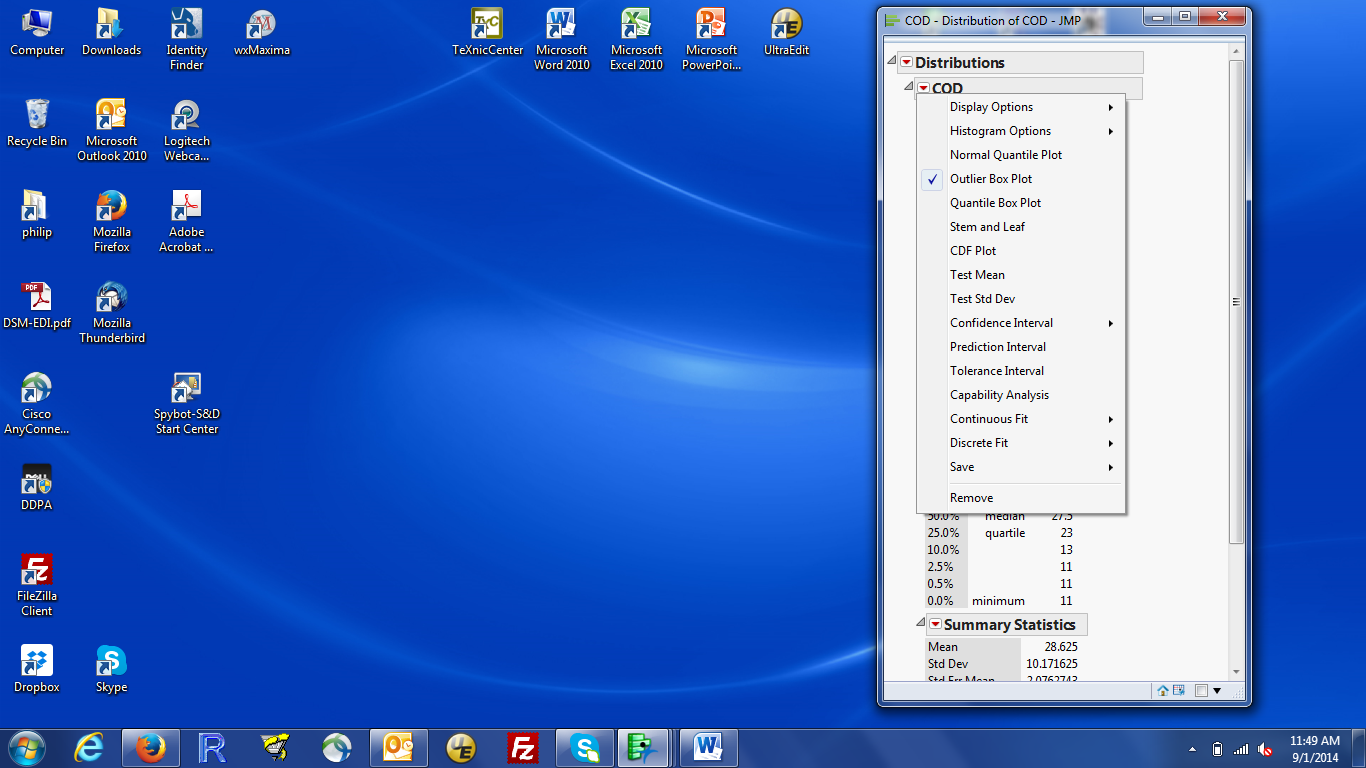
Make sure you know how these numbers are used (see lecture notes or the book), except perhaps for confidence intervals, which we will talk about in lecture this week.



Notice that JMP adds a red bar and a diamond to the “usual” box plot. To learn more about them, visit JMP Help / Books / Basic Analysis and look for Outlier box plot in the index

5. Notice that there is no menu bar at the top of the Distributions output window. There is one, but it is hidden until you look for it. Hover the mouse over the faint blue bar (not the darker blue window title). The bar will darken, then a menu will appear. If the bar darkens without a menu, left click on the dark blue bar. The menu will appear.

1. The 6 summary statistics shown by JMP are the default choice. Many more are available by clicking the red triangle by Summary Statistics then clicking “Customize Summary Statistics”. One especially useful choice is at the very bottom of that dialog: Enter (1-alpha) for mean confidence interval. The default is 0.95 for a 95% confidence interval. To get a 90% interval, change to 0.90; to get a 99% interval, change to 0.99.
2. JMP can do many more calculations and draw many more types of plots. To access these, click the red triangle **by COD** orright-click on the variable name (COD). There are three red triangles (by Distributions, by Summary Statistics, and by COD). You need to click the one by COD. A dialog box will pop up looking like:



This provides another way to get confidence intervals other than the 95% CI.

The Test Mean option provides access to hypothesis tests about the population mean.

1. In lecture, we looked (or will shortly look) at the hypothesis test that the population mean equals 32.5. To compute this, click Test Mean, enter the population mean under the null hypothesis, and click OK.

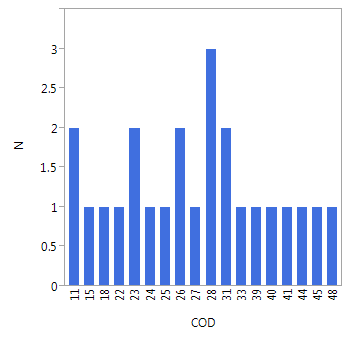
The two-sided p-value is given in the Prob > |t| line.

Note: JMP reports 3 p-values. Make sure you know which value is the two-sided p-value.

1. To get more summary statistics, left click the red triangle by Summary Statistics, then select Customize Summary Statistics. Select the statistics you want, then OK (very bottom of the window). JMP now reports the requested statistics.
2. When I produce a report that presents results from a statistical analysis, I find it easier to retype numeric results so I include only the results I need in my report. However, it is convenient to be able to copy graphs or plots from JMP without printing them.

You can copy and paste the contents of a window by left clicking in the window to make it active, using the mouse by activating the menu (see 5 above), then selecting edit / copy.

Or, use keyboard shortcut keys, which I find much easier once you remember the keys. Left click on the window, to make it active, then type ctrl-c (hold down the Ctrl key while typing once on the c key). That copies the graph to the clipboard. Navigate to your Word (or other word processing) document and type ctrl-v to paste the graph into your report. If you left click on the graph in Word, you activate the picture menu that allows you to resize or edit the graph. Here’s a copy and paste of the bar chart of COD values, using Graph/Chart.



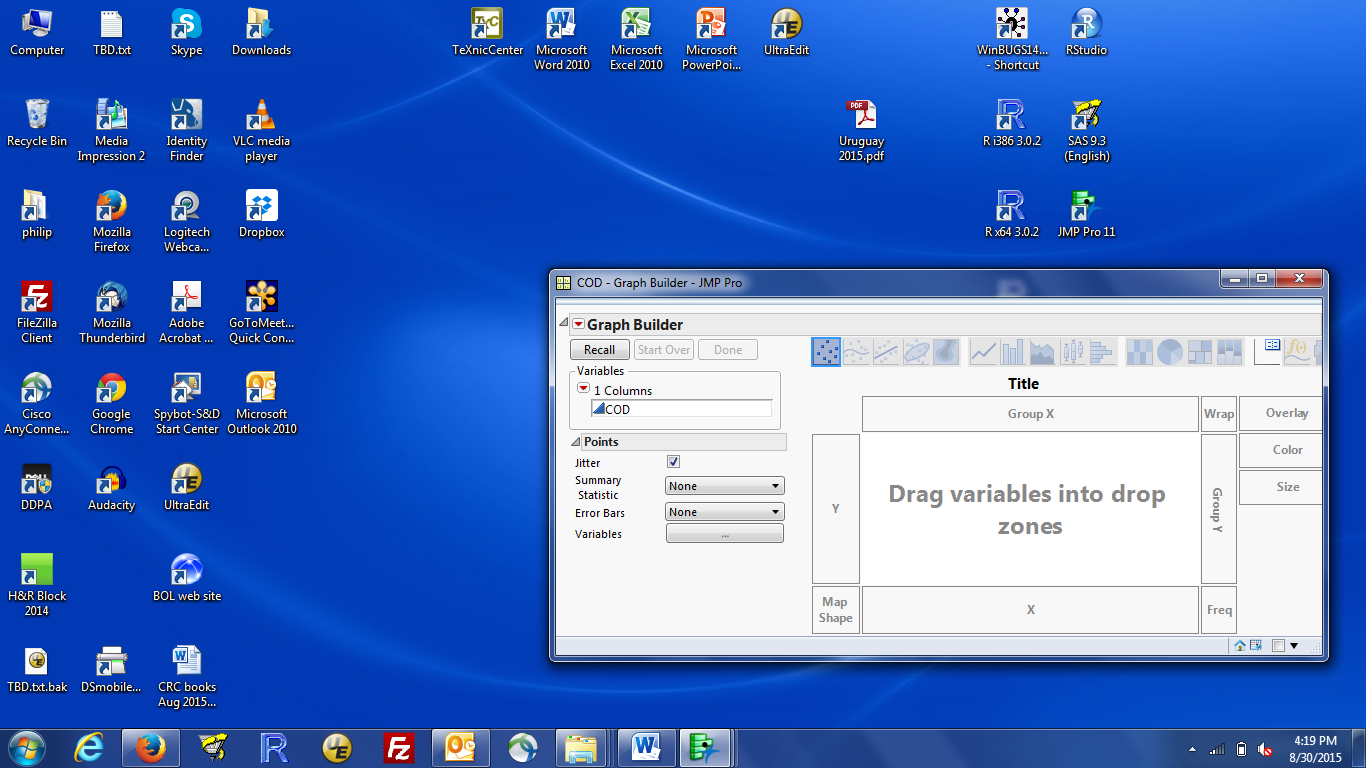
If you click on the window with the Analyze/Distribution results, you can copy the entire window into Word, then delete the parts you don’t want. Here’s the histogram and box plot for the COD data, after deleting the other pieces of the output.



If you want to get rid of the box plot, do that before copying the output. Click on the red triangle by the variable name (COD) and uncheck “Outlier Box Plot”. The box plot will disappear from the JMP output, so when you copy that output, you don’t get the box plot.

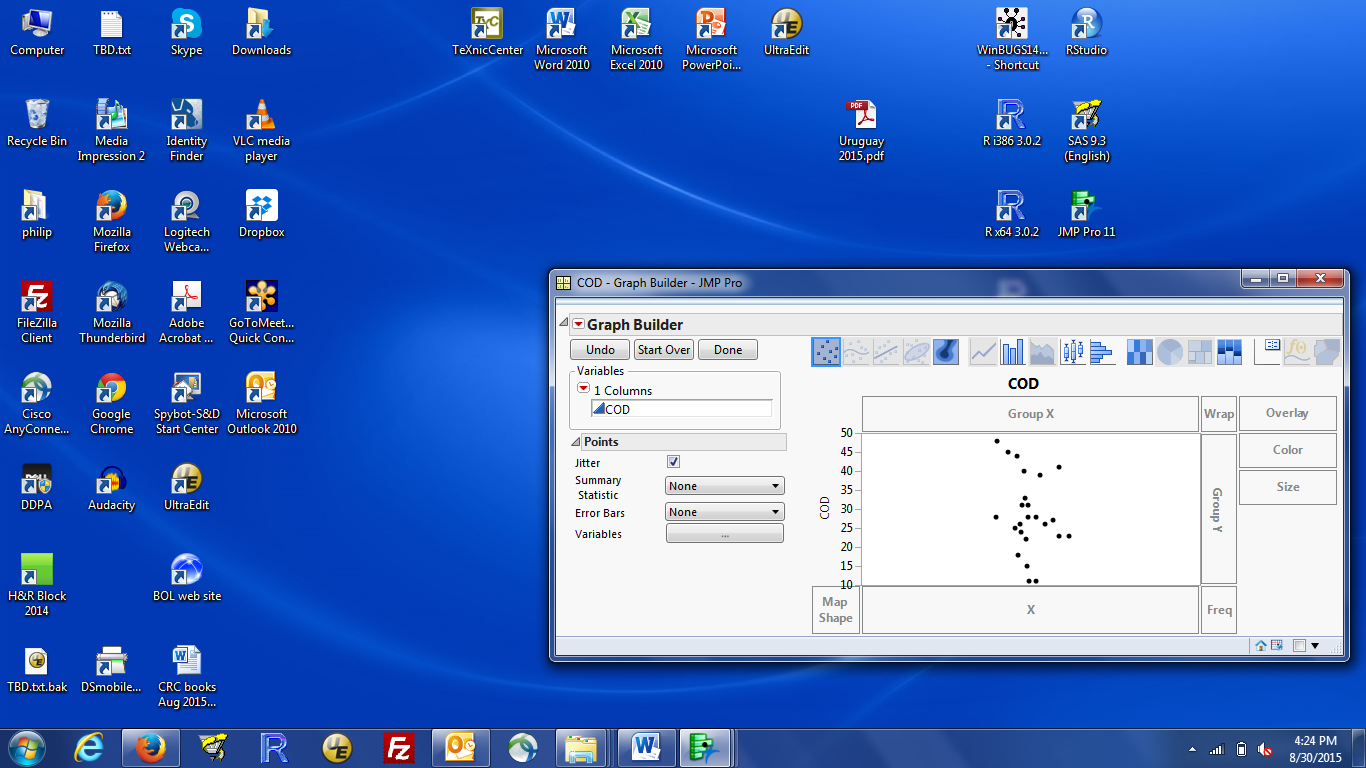
1. Often, you want a better box plot than the one produced by Analyze / Distribution. The Graph / Graph Builder platform provides the ability to draw all sorts of graphs.

Start the platform by Selecting Graph from the window with the COD data, then Graph Builder. You get an interactive display, with the variables in the data on the left.

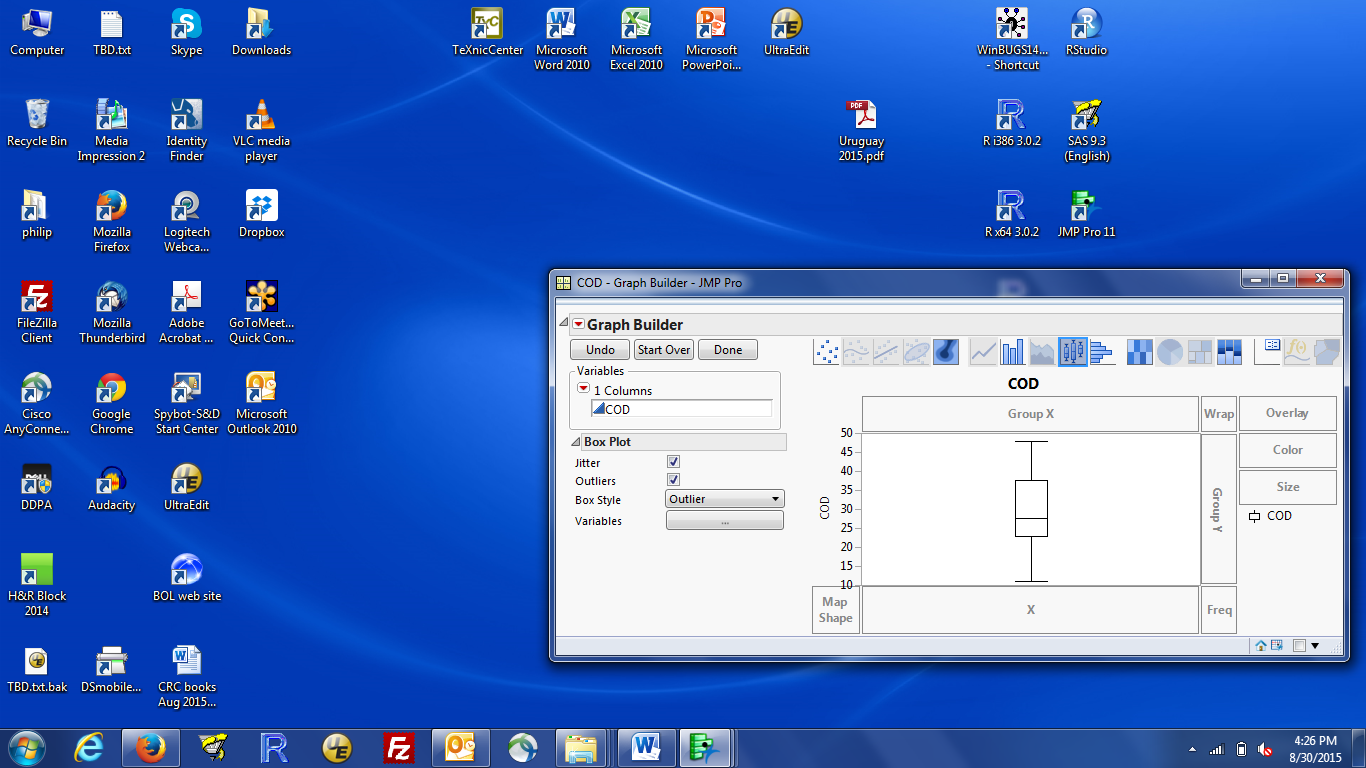


You build the graph you want by dragging variables into appropriate places in the righthand part of the window. The most frequently used places are Y and X.

To get a box plot of one group of observations, click the variable you want to plot in the Variables list then drag it to the Y area. You will see the COD values arranged vertically:



Turn this into a box plot by click on the boxplot icon at the top of the window (9th icon from the left). You should get:



Click the Done button (top left of the window) and you will get a stand-alone window with the plot. You can then copy this to clipboard (remember the hidden menu or keyboard shortcut).

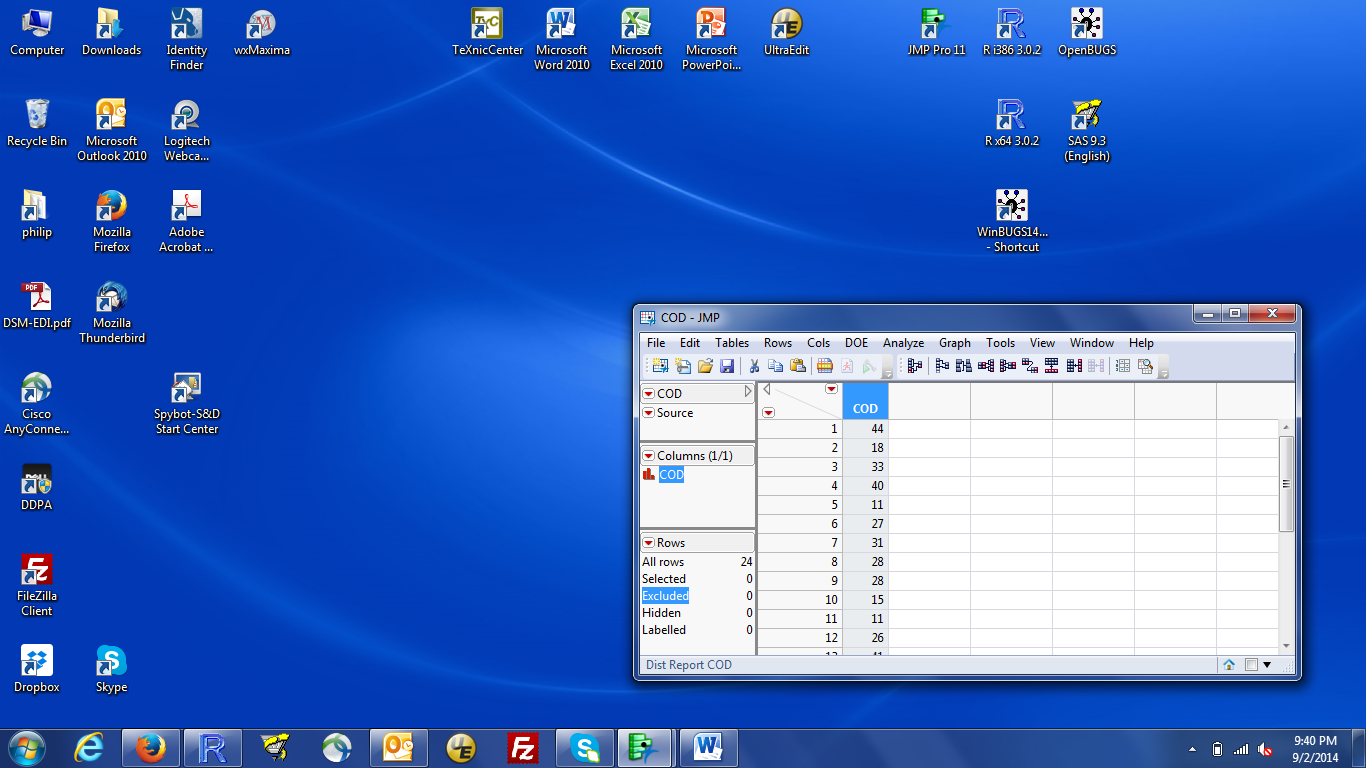
For fun, see what happens when you drag COD to the X variable area, then make a box plot. You could put the same variable in both X and Y areas, but that’s not a very useful plot or boxplot.

1. To draw side by side box plots, put the response variable into the Y variable area then put the grouping variable into the X variable area. You will (or should) see 2 or more columns of dots, one for each group. Click on the boxplot icon and you get side by side box plots.

If you did the ‘for fun in 11’, you can probably guess what happens if you put the response variable into the X variable area and the grouping variable into the Y area.

1. The final thing to demonstrate this week is the importance of the Modeling type. When you select Analyze/Distribution for a continuous variable, like COD, you get the histogram, quantiles, and numeric statistics (average, standard deviation, …).

Let’s see what happens when the Modeling type is incorrectly set. Go to the window with the COD data set and find the Columns box (on the left side, middle). There should be one entry, COD, because there is one variable, called COD, in the COD data set. There is a blue ramp (or triangle) next to the COD variable name as a visual indication that COD is a quantitative variable. Either select Cols / Column Info from the window menu or click on the blue ramp by COD. Change the modeling type from Continuous to Nominal (qualitative variable). The blue ramp should change to a red bar chart-like visual (see below, left middle):



Now select Analyze/Distribution. The output is now very different: a bar chart and list of frequencies (see output below). And, the options available by clicking on the red triangle are completely different. Almost certainly not what you wanted.

If JMP doesn’t behave as expected or doesn’t give you the options you expected, check that the modeling type is correct.

**Distributions**

**COD**



**Frequencies**

| **Level** | **Count** | **Prob** |
| --- | --- | --- |
| 11 | 2 | 0.08333 |
| 15 | 1 | 0.04167 |
| 18 | 1 | 0.04167 |
| 22 | 1 | 0.04167 |
| 23 | 2 | 0.08333 |
| 24 | 1 | 0.04167 |
| 25 | 1 | 0.04167 |
| 26 | 2 | 0.08333 |
| (Rest of the rows deleted) | | |
| Total | 24 | 1.00000 |

**Self Assessment:**

The data in wine.csv are the per capita annual wine consumption for 18 countries (units are liters per person per year). For the questions below, treat the data as if they were a simple random sample of 1st-world countries.

1) Draw a boxplot of the data. Is the data symmetric or skewed?

2) Estimate the mean per capita annual wine consumption in 1st world countries. Make sure to include units.

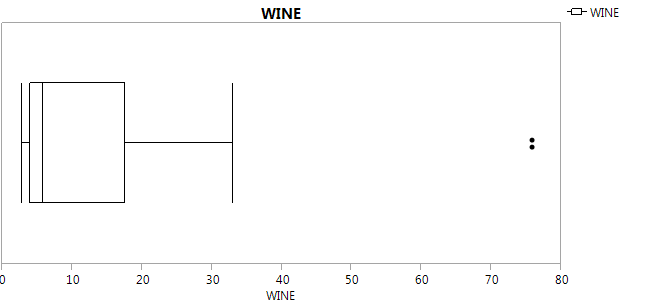
3) What are the standard deviation and standard error for wine consumption.

4) What is the variance for wine consumption (use JMP, not a calculator)

5) If you wanted to describe the variability in wine consumptions among 1st world countries, which number would you report: 23.1 or 5.4? Briefly explain your answer.

**Answers:**

1) Here’s the horizontal box plot:



These data are very skewed.

2) 16.5 liters/person year.

3) 23.1 liters/person year and 5.4 liters/person year

4) 533 squared liters/person year.

5) You should report 23.1 (the sd). That’s because the question asks about variability among the observations.