

INTRODUCTION TO MYSTAT

Go to the course Blackboard site and download “Laboratory 3 MYSTAT Intro.xls” – open this file in Excel. This file contains data on a watermelon seed spitting contest wherein 3 teams spit watermelon seeds of varying weights (in mg) from different postures (either standing = “high” or reclining = “low”) and either “against” the wind or “with” the wind. The Posture and Wind variables were also combined into a single “Treatment” variable for making some analyses easier. The distance that each seed was spit is shown in cm.

Open Mymstat on your computer.

The pages or tabs you want active at the top left of the screen in Mymstat are “Untitled.syz” and “Untitled1.syz*” – close other pages if they’re open for now if you want the workspace less cluttered.

Click on the tab at the top of the window titled “Untitled1.syz*”

This spreadsheet is the input for Mymstat, it is where you will enter data directly into or copy and paste from an Excel spreadsheet.

Click on the second tab at the top of the window titled “Untitled.syz”

This is where Mymstat dumps the output from any statistics generated from your data.

Go back to the input spreadsheet.

Data Entry

Entering data into Mymstat is very much like Excel except each column is called a “Variable” and each variable must be defined. This is different from Excel in that the heading of each column cannot be located in “row 1,” but you must rename the column headings which are right now denoted as just “VAR.”

Rename each variable (column) in Mymstat to match your column names in Excel.

Do this by right clicking on the first variable in Mymstat and selecting “Variable Properties.”

For “Variable name” enter the first column heading from your Excel spreadsheet data.

(Mymstat cannot handle spaces between any characters - to enter compound variable names headings you must separate the words by underscores “_” or do not include a space between them – keep variable name as simple as possible.

Any data which contain alphabetic characters anywhere must be classified as a “String” variable (e.g., the “posture” and “wind” variables); under “Variable type” select “String.” Because these data are not purely numeric you will also have to insert a dollar sign “\$” after the variable name. This is because every variable that does not have a \$ after it is considered by Mymstat as numeric.

Either numeric or string data can be “categorical,” and that should be indicated as such under “Variable type.”

Select “OK” and rename the remaining variables as you just did to match the column headings from the Excel spreadsheet. Make sure to select “Numeric” or “String” under “Variable type” appropriately and add the \$ to variable names as necessary. Also, in the same window you can change the number of decimal places shown.

Next, copy and paste the data from the Excel spreadsheet into the Mypstat spreadsheet. To do this highlight all of your data in Excel, excluding the headings for each column, right click and select “Copy.”

Go back to the Mypstat input spreadsheet that you have prepared and right click in the first cell under your first variable heading. Select “Paste.”

All of your data from Excel should now be present in Mypstat, but you should always double check by scrolling down making sure all the data were transferred correctly, all the way to the last observations (make sure there are as many rows in Mypstat as in Excel).

Save the data file to your disk – use a descriptive file name so that you remember what these data are for, and notice that the filename is given the suffix “.syz” which is a Mypstat data file. The tab has now also changed to reflect your new filename so that you can always keep track of which data you’re working on.

Transforms

Data transformation is fairly straightforward in Mypstat. Click on the “Data” menu at the top of the window and select “Transform” then “Let..” Type “log_weight” in the “Let” box. Then click inside the “expression” box, double click “LOG” in the “Functions” box, and double click “Distance” in the “variable(s)” box. Notice that you’ve set up a little algebraic equation that will create a new variable which is the natural logarithm (ln) of Distance. Click “OK” and the new variable will appear, properly named, in the seventh variable column.

Now, ln-transform Weight and add it in as a new column, and then create another new variable which is Distance divided by Weight (the “expressions” box recognizes the usual mathematical operators). You should now have a total of nine variables in the dataset – be sure to save your work.

Play with the transforms if you want to... notice that there lots of options in the “Functions” box (most of which, I will freely admit, I know nothing about!). Notice, however, that transforming some kinds of variables is not meaningful.

Basic Statistics

Next run some basic statistics on the data set such as mean, range, and variance.

Click on “Analyze” at the top of the window and select “Basic Statistics”

Variables in the dataset are shown in the “Available variables” window. Click on a variable that you want to analyze, and then click “Add” to move it to the “Selected variable” window. Select as many variables as you want. Keep in mind that getting descriptive statistics for some of the variables is not very useful (e.g., why would you want descriptive statistics for the “Group” variable??).

Under “Options,” check the boxes for the descriptive statistics that you want returned for each variable that you selected above. Make sure that N, Minimum, Maximum, Arithmetic mean, SE of AM, SD, CV, Range, and Variance are all selected. Do not worry about the choices below the “Options” window for now.

Select “OK.” This will then take you to the Output screen (“Untitled.syo”) with the results of the statistics you chose to run for each variable. All of the statistics requested for each variable are shown in a table – this may be a much faster way to obtain descriptive statistics compared to using Excel (with the nasty exception that you can’t copy-and-paste from this window!). You can print the Output window, and you can save the output of your analyses into a file that will have the suffix “.syo” (but as far as I can tell, these can only be opened in Mstat).

You can also obtain descriptive statistics on your data for groups that you define separately. Click on “Data” at the top of the window and select “By Groups.” Click on “Team” and then “Add” and then “OK.” Now ask again for the descriptive statistics on the same variables you specified above (although make sure that “Team” is not one of the variables being analyzed!).

Notice that you now get descriptive statistics for each team separately.

Go back to “By Groups,” “Remove” Team as the grouping variable, and replace it with “Treatment,” then go and get the descriptive statistics. Notice that all stats are presented for each treatment group separately.

Now – go back to “By Groups,” leave “Treatment” in the “Selected variables” box, but add “Team” back as a grouping variable (i.e., you should have both Treatment and Team in the box). Go back and ask for the descriptive statistics again – note that you now have all the descriptive statistics for each Treatment-by-Team combination separately. Again, this can be a much faster way to obtain these results than using Excel.

NOTE: Be sure to pay attention to the status of the By Groups menu if and when you start using it. If you call for a procedure without realizing that By Groups is on or how it is grouping output, you can get some weird results!

Graphing

Bar chart (remember: used when X-axis variable is not continuous variable): you can either click the “Graph” menu and select “bar chart,” or click on the little tiny bar chart icon on the toolbar. Be sure By Groups is turned off/emptied of variables.

Select Treatment as the X-variable, Distance as the Y-variable, and select “Standard error” under the “Error Bars” tab and click OK. The bar chart with error bars should pop up, both on a new page labeled “Graph1” and on the output page.

Note that there are a variety of options that you can change on the graphs in the Bar Graph window. You can copy-and-paste graphs from the Graph page, but not from the output page; you can also save a graph to a variety of formats to bring into other programs. There are additional editing options on the Graph page, but recognize that it may be easier to format graphs to guidelines from Lab 2 using Excel rather than Mystat.

The above bar graph showed you mean Distance traveled (\pm SE) for each of the four Treatments, across all three Teams together. To see mean Distance traveled for each Treatment *within* each Team, you can either (a) select Team in the By Groups menu, as above or (b) add Team to the “Grouping variable” list on the Bar Chart menu when you first make the bar chart, then make the bar chart as above. Notice that the Graph page shows only the last graph made; all three graphs appear in the output window, but you can’t copy-and-paste from the output page.

Quick side note: If your Output page is getting cluttered with lots of output such that it’s hard for you to find stuff you completed earlier easily, you can clear the output page by right-clicking anywhere on the output page, and selecting “Clear output” (you could also open a new output page [here](#)).

Scatterplot (Be sure By Groups is turned off and the selection window is empty).

Click on the Graph menu and select “Scatterplot,” or click the small Scatterplot icon on the toolbar.

While under the “Main” tab, click on “Weight” and add it as the X-variable (this is the independent variable), then click “Distance” and add it as the Y-variable (the dependent variable). Under the “Smoother” tab, select “Linear” (this will add the “best fit line” to your scatterplot to help visualize any trend that might be present. Click OK. The scatterplot should show up on both the Graph page and the Output page.

Change the dependent variable from Distance to the Distance-divided-by-Weight variable that you created at the beginning of this exercise. (Quite a different looking plot, yes??)

Go back and create a Distance by Weight scatterplot (Distance as Y-axis and Weight as X-axis) for each Team separately (you can use the same strategies as under Bar Chart: select Team as a By Group variable, or add Team as the Grouping variable on the Scatterplot creation window). Note that all three scatterplots come up on the Graph page and on the output page.

Now create a Distance by Weight scatterplot for each Treatment *within* each Team (hint: you can add more than one grouping variable to the By Groups list, and you can add more than one variable to the Grouping variable pane in the Scatterplot creation window). You should see a total of 12 graphs pop up (4 treatments within each of 3 teams).

Now create a Distance by Weight scatterplot for the whole dataset (turn off all grouping options). Under the “Symbol and Label” tab, in the “Symbol Type” subwindow, select Treatment in the “Select Variable” pull down menu. Click OK. You should now see the scatterplot you saw in the first step of this exercise, but with the data points coded as to which Treatment they are in.

This allows you to see if there is any systematic bias in your data points not immediately obvious in other plots.

Finally, create a Distance by Team scatterplot (Distance as Y, Team as X), without using any grouping variables (i.e., turn off By Groups and/or make sure the Grouping variable pane is empty). This isn't really a scatterplot per se, but this is kind of a nifty way to see how individual observations looked in each team in your whole dataset.

And Finally (no, really).... Weighted Means and Variances

Enter the following dataset into a new Mypstat data sheet:

<u>Frequency</u>	<u># of males</u>	<u>Condition</u>
1	0	study1
8	1	study1
75	2	study1
169	3	study1
62	4	study1
8	5	study1
2	6	study1
5.1	0	expected
30.5	1	expected
76.2	2	expected
101.6	3	expected
76.2	4	expected
30.5	5	expected
5.1	6	expected
20	0	study2
40	1	study2
58	2	study2
86	3	study2
55	4	study2
43	5	study2
23	6	study2

Shown are the frequency distributions for the sex ratio studies (Study #1 and Study #2) from Monday's lecture. To find the weighted mean and variance in the number of male offspring, you need to tell Mypstat that the "frequency" variable is the weighting variable (i.e., the frequency with which each # of males was observed). Do this by clicking on "Data," then "Case Weighting," then "By Frequency." Add your frequency variable name to the Selected Variables box and click OK.

Now find N, the mean, and the variance in the number of males for each Condition in the data set.... (you might take note of how the patterns reflect what we discuss in lecture....)