

Our objective for today is to examine the SAS statistical analysis software. Open SAS and familiarize yourself with the window components (Program, Log and Output). Then, write a program in the “program” window, run the program and work out any errors. The final product should be emailed to me.

On the following page is the program I showed in class (the larger one). The sections in bold represent the minimal program needed to accomplish the assignment. Add to this a comment with your name. Feel free to explore as many additional instructions and options as you wish.

We will have available both an ASCII (TXT) files that can be used as a part of the program and CSV files that are better as external files. Today you may wish to download the TXT version of data set Ex0428.txt and save it in the C:\TEMP directory. You may then change the TXT to SAS and build your SAS program around the dataset.

If you prefer, you can use the “filename” and “infile” statements to reference an external file. In the future we will call in an external file, but this will not be required today. When using external files, temporarily store files in the “C:\TEMP\” directory. Also, if you wish to output an HTML file, store it in the “C:\TEMP\” directory. These directories will be erased when you log off the machine, so anything you want permanently should be placed in Tiger bytes.

Darwin data: The data represents seeds from flowers of 15 plants. Half of the flowers were pollinated with pollen from the same plant (self fertilized) and half were pollinated with pollen from distant plants (cross fertilized). Both sets of seeds were germinated and one seedling from each plant and pollination treatment was grown for a given amount of time and the heights taken. The data set contains these plant heights in two columns (Cross and Self respectively).

23.5	17.375
12	20.375
21	20
22	20
19.125	18.375
21.5	18.625
22.125	18.625
20.375	15.25
18.25	16.5
21.625	18
23.25	16.25
21	18
22.125	12.75
23	15.5
12	18

For your program treat the cross fertilized values as the dependent variable (Y variable ) and the self fertilized values as the independent variable (X variable). Using these variables do the following.

- List the data with PROC PRINT
- Get a scatter plot of the data with PROC PLOT
- Run PROC UNIVARIATE on the data (variable CROSS only)
- Run a regression on the data using PROC REG